**Remarks**

The Office Action dated April 7, 2006 and Advisory Action dated August 1, 2006 have been reviewed, and the following remarks are made in response thereto. In view of the following remarks, Applicants respectfully request reconsideration of this application and timely allowance of the pending claims. Upon entry of the instant amendment, claims 1, 42 and 103-109 are pending. Claims 1 and 42 have been amended. Written support for the claim amendments is found throughout the specification and in the original claims, thus Applicants submit that no prohibited new matter has been added.

The previous amendment and response filed by Applicants on May 31, 2006 was purported to be non-compliant for failing to properly specify changes made to the specification. Applicants respectfully submit that this amendment and response indicates changes made to the specification in compliance with 37 C.F.R. 1.121.

Moreover, Applicants submit that this Amendment should be entered because the amendments place the application in better condition for appeal. In particular, the specification incorporates by reference several scientific journals that contain the full name for the abbreviated terms GST, SH1, PH and PTB (paragraph that begins on page 31, line 22 of the specification). Thus, entry of the full name for each abbreviated term into the specification does not constitute new matter.

Summary of the final Office Action

1. The previous amendment and response filed by Applicants on May 31, 2006 was purported by the Examiner to be non-compliant.
2. A new oath/declaration was requested.
3. Claims 1, 42 and 103-109 are rejected under 35 U.S.C. 112 (second paragraph) as being indefinite because the abbreviations used in claims 1 and 42 are subject to more than one interpretation.
4. Claims 1, 42 and 103-109 are rejected under 35 U.S.C. 112 (second paragraph) as being unclear because it is uncertain if "GST" as used in claims 1 and 42 refers to a whole protein or a domain of a protein.

Summary of the Advisory Action

1. The Examiner refused to enter the after final amendment because there does not exist support in the specification as filed for the proposed claims amendments which indicate the full name for each of the abbreviated protein domains.

Notice of Defective Oath/Declaration

The Examiner requested a new oath/declaration because the declaration allegedly contained non-initialed and/or non-dated alterations. Applicants respectfully traverse the objection.

Applicants would like to thank the Examiner for the phone conversation held with Applicants' Agent on July 6, 2006. During this conversation, the Examiner agreed that the supplemental declaration submitted on June 24, 2004 corrected the deficiencies with respect to Andrew Sparks in the original declaration. Briefly, Applicants submitted that Andrew Sparks properly executed a supplemental declaration in which he was the sole signatory. Under 37 C.F.R. 1.67(a)(2) deficiencies or inaccuracies relating to fewer than all the inventors may be corrected with a supplemental oath/declaration identifying the inventive entity but signed only by the inventor(s) to whom the error or deficiency relates. Andrew Sparks properly executed a supplemental declaration, without any alterations, that was filed simultaneously with the original (allegedly defective) declaration on June 24, 2004. Therefore, Applicants respectfully request that the Examiner withdraw the notice of defective oath/declaration.

Rejections under 35 U.S.C. 112 (Second Paragraph)

Claims 1, 42 and 103-109 were rejected under 35 U.S.C. 112 (second paragraph) as being indefinite because the acronyms "GST, SH1, PH, PTB and LIM," in claims 1 and 42 are allegedly subject to more than one interpretation. Applicants respectfully traverse the rejection.

Applicants submit that under MPEP 2111.01, "[T]he ordinary and customary meaning of a claim term is the meaning that the term would have to a person of ordinary skill in the art in question at the time of the invention, *i.e.*, as of the effective filing date of the patent application." Applicants respectfully submit that claim 1 reads in part, "a domain of interest selected from the group consisting of GST, SH1, PH, PTB, [and] LIM." Given that the abbreviated terms "GST, SH1, PH, PTB and LIM" are used in combination with the preceding term "domain," these terms would be well understood and clearly defined to the skilled artisan. Accordingly, Applicants respectfully submit that the skilled artisan would have understood the acronyms that were used in the pending claims at the time the instant application was filed. Specifically, Applicants provide herein several journal articles that clearly define the acronyms GST, SH1, PH, PTB and LIM when used in the context of a protein domain. In particular, Cohen *et al.* (1995) Cell 80:237-248 and Pawson *et al.* (1995) Nature 373:573-580 provide a full name for PH and Src domains, Michelsen *et al.* (1993) PNAS USA 90:4404-4408 provides the derivation of LIM domain,

Toung *et al.* (1990) PNAS USA 87:31-35 provides the full name for GST used in a protein context and Van Der Geer *et al.* (1996) PNAS USA 93:963-968 provides the full name for PTB. As such, it is apparent that the skilled artisan could readily access the meaning of the acronyms in the claims at the time the application was filed. Therefore, Applicants respectfully request reconsideration and withdrawal of the rejection of claims 1, 42 and 103-109 under 35 U.S.C. 112 (second paragraph).

Even assuming *arguendo* that the abbreviated claim terms are subject to more than reasonable interpretation, without acquiescing to the merits of the Examiner's rejection, Applicants have amended the claims to clearly indicate the specific full-name for each abbreviated term. Specifically, Applicants amended claims 1 and 42 to specify that the term "GST" refers to Glutathione S-transferase, the term "SH1" refers to Src homology 1, the term "PH" refers to Pleckstrin homology and the term "PTB" refers to Phosphotyrosine binding. Concurrently, Applicants have amended the specification to include the full-name for each abbreviated term in order to provide clear antecedent basis for the full-names used in the amended claims. While an Applicant is not limited to the nomenclature used in the application as filed, he or she should make appropriate amendment of the specification whenever this nomenclature is departed from by amendment of the claims so as to have clear support or antecedent basis in the specification for the new terms appearing in the claims (see MPEP 608.01(o)). Given that the specification incorporates by reference several scientific articles (*e.g.* Cohen *et al.* and Pawson *et al.*), the full name for the abbreviated claim terms are supported by the specification. Accordingly, Applicants respectfully request that the rejection under 35 U.S.C. 112 (second paragraph) be reconsidered and withdrawn.

Conclusion

The foregoing amendments and remarks are being made to place the application in condition for allowance. Applicants respectfully request entry of the amendments, reconsideration, and the timely allowance of the pending claims. A favorable action is awaited. Should the Examiner find that an interview would be helpful to further prosecution of this application, they are invited to telephone the undersigned at their convenience.

If there are any additional fees due in connection with the filing of this response, please charge the fees to our Deposit Account No. 50-0310. If a fee is required for an extension of time under 37 C.F.R. 1.136 not accounted for above, such an extension is requested and the fee should also be charged to our Deposit Account.

Respectfully submitted,
**Morgan, Lewis & Bockius LLP**

R. Smyth RN: 50801

Robert Smyth for

Christopher J. Betti, Ph.D.
Reg. No. 56,890

Dated: **October 10, 2006**
Morgan, Lewis & Bockius LLP
Customer No. **09629**
1111 Pennsylvania Avenue, N.W.
Washington, D.C. 20004
202-739-3000

Proc. Natl. Acad. Sci. USA
Vol. 93, pp. 963–968, February 1996
Biochemistry

# Identification of residues that control specific binding of the Shc phosphotyrosine-binding domain to phosphotyrosine sites

(tyrosine phosphorylation/signal transduction)

PETER VAN DER GEER*†, SANDRA WILEY*†, GERALD D. GISH*†, VENUS KA-MAN LAI*†, ROBERT STEPHENS‡, MORRIS F. WHITE§, DAVID KAPLAN‡, AND TONY PAWSON*†¶

*Programme in Molecular Biology and Cancer, Samuel Lunenfeld Research Institute, and †Protein Engineering Network Centres of Excellence, Mount Sinai Hospital, 600 University Avenue, Toronto, ON Canada M5G 1X5; ‡Advanced Bioscience Laboratory–Basic Research Program, National Cancer Institute–Frederick Cancer Research and Development Center, Frederick, MD 21702; and §Joslin Diabetes Center and Department of Medicine, Brigham and Women's Hospital and Harvard Medical School, Boston, MA 02115

*Communicated by Henry G. Friesen, Medical Research Council of Canada, Ottawa, Canada, November 6, 1995 (received for review August 29, 1995)*

ABSTRACT **The Shc adaptor protein contains two phosphotyrosine [Tyr(*P*)] binding modules—an N-terminal Tyr(*P*) binding (PTB) domain and a C-terminal Src homology 2 (SH2) domain. We have compared the ability of the Shc PTB domain to bind the receptors for nerve growth factor and insulin, both of which contain juxtamembrane Asn-Pro-Xaa-Tyr(*P*) motifs implicated in PTB binding. The Shc PTB domain binds with high affinity to a phosphopeptide corresponding to the nerve growth factor receptor Tyr-490 autophosphorylation site. Analysis of individual residues within this motif indicates that the Asn at position −3 [with respect to Tyr(*P*)], in addition to Tyr(*P*), is critical for PTB binding, while the Pro at position −2 plays a less significant role. A hydrophobic amino acid 5 residues N-terminal to the Tyr(*P*) is also essential for high-affinity binding. In contrast, the Shc PTB domain does not bind stably to the Asn-Pro-Xaa-Tyr(*P*) site at Tyr-960 in the activated insulin receptor, which has a polar residue (Ser) at position −5. Substitution of this Ser at position −5 with Ile markedly increased binding of the insulin receptor Tyr-960 phosphopeptide to the PTB domain. These results suggest that while the Shc PTB domain recognizes a core sequence of Asn-Pro-Xaa-Tyr(*P*), its binding affinity is modulated by more N-terminal residues in the ligand, which therefore contribute to the specificity of PTB–receptor interactions. An analysis of residues in the Shc PTB domain required for binding to Tyr(*P*) sites identified a specific and evolutionarily conserved Arg (Arg-175) that is uniquely important for ligand binding and is potentially involved in Tyr(*P*) recognition.**

Shc is a member of a group of cytoplasmic signaling proteins we have collectively termed adaptor proteins (1). Shc proteins become phosphorylated on tyrosine after stimulation with a wide variety of growth factors and cytokines (2–10). Phosphorylation of mammalian Shc on Tyr-317 creates a binding site for the SH2 domain of Grb2, which is in turn associated with the Ras guanine nucleotide exchange factor Sos (11). The interaction of Shc with the Grb2–Sos complex may provide a mechanism for Ras activation (3, 12–17).

Shc contains an N-terminal phosphotyrosine [Tyr(*P*)] binding (PTB) domain, a central Gly/Pro-rich region that contains the principal Tyr phosphorylation site at Tyr-317, and an SH2 domain at its C terminus. The Shc SH2 domain recognizes Tyr(*P*) in the context of C-terminal residues and binds preferentially to phosphopeptides with the sequence pY-E/L/I/Y-X-L/I/M, where pY is Tyr(*P*) (18). The PTB domain, which is highly conserved in Shc-related proteins, was recently identified as a sequence of ≈160 amino acids that bind specific Tyr(*P*)-containing proteins (19–22). Both the sequence and binding properties of the Shc PTB domain appear quite different from those of known SH2 domains.

Shc binds to activated growth factor receptors through the recognition of phosphorylated Tyr residues that in many cases lie within the sequence NPXpY (23, 24). The Shc PTB domain binds to the activated nerve growth factor receptor (NGFR) and the activated epidermal growth factor receptor (EGFR), which possess such NPXpY motifs, and to a 145-kDa protein that becomes phosphorylated on Tyr in platelet-derived growth factor-stimulated cells (19–21, 25). The NGFR binds Shc through a single juxtamembrane autophosphorylation site, Tyr-490, that lies within an NPXY element (23, 26). A mutant NGFR with Phe in place of Tyr-490 fails to interact with Shc *in vivo* or with the Shc PTB domain *in vitro* (19, 23).

These results suggest that the Shc PTB domain recognizes Tyr(*P*) residues within the sequence NPXpY and, therefore, differs from SH2 domains in the sense that it recognizes specific residues N terminal to the Tyr(*P*). Consistent with this view, the Shc PTB domain recognizes specific phosphopeptides with such motifs *in vitro* (25, 27–30). Here we have investigated the residues in both the PTB domain and its ligands that control complex formation.

## MATERIALS AND METHODS

**Cell Lines, Antisera, and Fusion Proteins.** CHO cells expressing wild-type (wt) insulin receptors (IR) (31) were grown in F-12 medium containing 25 mM Hepes (pH 7.4) and 10% (vol/vol) fetal bovine serum. NIH 3T3 cells expressing wt and Phe-490 mutant NGFRs (23) were grown in Dulbecco–Vogt's modified Eagle's medium (DMEM) containing 10% (vol/vol) calf serum. NIH 3T3 cells overexpressing the human EGFR (32) were grown in DMEM containing 10% calf serum and G418 (400 μg/ml). The anti-IR monoclonal antibody 51 was obtained from I. Goldfine (33, 34). Polyclonal anti-NGFR and anti-Shc sera have been described (6, 35). The anti-Tyr(*P*) monoclonal antibody 4G10 was obtained from Upstate Biotechnology (Lake Placid, NY). The glutathione *S*-transferase (GST)–Shc PTB fusion protein used for receptor binding experiments is identical to GST–ShcB as described (19). The GST–dShc PTB fusion protein has been described (22).

**Immunoprecipitations and PTB Binding Assays.** Cells were grown to confluence and starved 16 hr in serum-free medium. CHO cells expressing the IR were stimulated with 100 nM

The publication costs of this article were defrayed in part by page charge payment. This article must therefore be hereby marked "*advertisement*" in accordance with 18 U.S.C. §1734 solely to indicate this fact.

Abbreviations: Tyr(*P*) or pY, phosphotyrosine; PTB, Tyr(*P*) binding; NGFR, nerve growth factor receptor; EGFR, epidermal growth factor receptor; wt, wild type; IR, insulin receptor; GST, glutathione *S*-transferase; DER, *Drosophila* EGFR homologue.

¶To whom reprint requests should be addressed.

insulin for 5 min at 37°C. NIH 3T3 cells expressing NGFRs were stimulated with NGF (50 ng/ml) for 5 min at 37°C, and NIH 3T3 cells expressing the human EGFR were stimulated with EGF (100 ng/ml) for 5 min at 37°C. Cell lysates were made, and immunoprecipitations and PTB binding assays (in the absence or presence of 2 or 5 $\mu$M competing phosphopeptide) were performed as described (19).

**Surface Plasmon Resonance Analysis of Phosphopeptides Interacting with the Shc PTB Domain.** Peptides were synthesized using 9-fluorenylmethoxycarbonyl (Fmoc) solid-phase chemistry with direct incorporation of Tyr(*P*) as the $N^{\alpha}$-fluorenylmethoxycarbonyl-*O*-dimethylphosphono-L-tyrosine derivative. Cleavage of the peptide from the resin and deprotection was achieved through an 8-hr incubation at 4°C in trifluoroacetic acid containing 2 M bromotrimethyl silane and a scavenger mixture composed of thioanisole, *m*-cresol, and 1,2-ethanedithiol (1.0:0.5:0.1, percent by volume). The product was precipitated with ice-cold *t*-butyl ethyl ether and collected by centrifugation. After desalting of the crude material, pure phosphopeptide was isolated by using reverse-phase HPLC. The authenticity of the phosphopeptide was confirmed by amino acid analysis and mass spectroscopy. Surface plasmon resonance analysis was carried out using a Biacore apparatus (Pharmacia Biosensor) as described (36). The peptide LSLLSNPTpYSVMRSK was immobilized to a biosensor chip through injection of 0.5 mM phosphopeptide, in 50 mM Hepes, pH 7.5/2 M NaCl, across the chip surface previously activated by procedures outlined by the manufacturer. Typically in surface plasmon resonance experiments, a signal of 1500 resonance units was obtained from injection of GST–Shc PTB in the absence of competing peptide.

**Expression of Torso–*Drosophila* EGFR Homologue (DER) fusion protein in Transgenic Flies.** Transgenic flies expressing the activated Torso–DER chimeric protein, containing the cytoplasmic domain of DER, under the control of the heat shock promoter were obtained from E. Hafen (Universität Zürich, Zurich). Protein expression was induced by growing the flies at 37°C for 45 min after which they were allowed to recover at room temperature for 2.5 hr. Lysates were made as described (22).

## RESULTS

**Characterization of the Core PTB Binding Motif.** To investigate the contribution of the Asn and Pro residues within the consensus PTB binding site to phosphopeptide recognition, we changed these residues to Ala in a phosphopeptide based on the sequence around Tyr-490, the Shc-binding site in the NGFR. The wt and mutant peptides were tested for their ability to compete with NGFRs, present in lysates of NGF-stimulated cells, for binding to a GST fusion protein containing the Shc PTB domain (Fig. 1). The wt phosphopeptide (HIIENPQpYFSD) competed efficiently for binding. Changing the Asn at position −3 [relative to the Tyr(*P*)] to Ala completely abolished detectable binding to the PTB domain, as measured in the competition assay, whereas changing the Pro at position −2 to Ala induced a significant but more modest reduction in the affinity of the PTB–peptide interaction. To determine the contribution of the different residues in the NPXpY motif more precisely, various concentrations of wt or mutant phosphopeptides were tested for their ability to inhibit binding of the Shc PTB domain to a Tyr(*P*)-containing peptide, from polyoma middle tumor antigen (19, 24, 37), immobilized on a Biacore chip (Table 1). This analysis confirmed that the Asn residue at position −3 is essential for binding to the PTB domain *in vitro*, while the Pro at position −2 is less important. These findings are in agreement with those of others (28, 29).

**PTB Recognition Sites Contain Residues Outside the NPXpY Motif That Regulate Binding Specificity.** Our previous

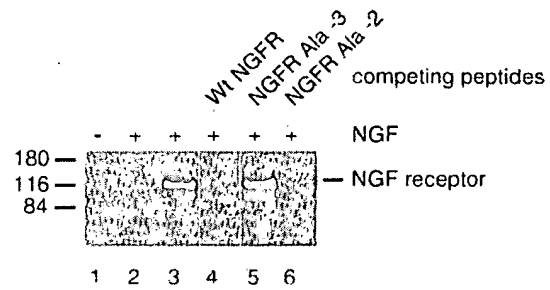



FIG. 1. Asn present within the NPXpY motif is essential for binding to the PTB domain. GST (lane 2) and GST Shc PTB (lanes 1 and 3–6) fusion proteins bound to glutathione-agarose were incubated with NGFRs present in lysates from control (lane 1) and NGF-stimulated (lanes 2–6) cells in the absence (lane 1–3) and presence of wt (lane 4) and mutant (lanes 5 and 6) competing Tyr(*P*)-containing peptides, based on the sequence around Tyr-490, the Shc PTB binding site in the NGFR. Bound proteins were examined by anti-Tyr(*P*) immunoblot analysis. Competing peptides: wt NGFR, HIIENPQpYFSD (lane 4); NGFR Ala −3 mutant, HIIEAPQpYFSD (lane 5); NGFR Pro −2 mutant, HIIENAQpYFSD (lane 6). Amino acid substitutions introduced in these peptides are shown in boldface type.

results have suggested that an NPXpY motif may be necessary but is not sufficient for stable PTB binding (19). To pursue the identification of residues that might modulate Shc PTB binding, we have employed two receptors, those for NGF and insulin, both of which have juxtamembrane NPXY autophosphorylation sites (Fig. 2*A*) but have markedly different abilities to interact with Shc. Shc coprecipitated with the wt NGFR from NGF-stimulated cells but not with a mutant receptor with a Phe replacing Tyr-490 (Fig. 2*B*). Two additional Tyr(*P*)-containing proteins of unknown identity also coprecipitated with Shc from lysates of control and NGF-stimulated NIH 3T3 cells expressing either wt or Phe-490 NGFRs (Fig. 2*B*, lanes 1, 2, 5, and 6). Consistent with previously published data, the NGFR bound to the Shc PTB domain *in vitro* in a fashion that was dependent on phosphorylation of the NGFR at Tyr-490 (Fig. 2*C*). In contrast to the NGFR, the activated IR failed to bind stably to Shc *in vivo* or *in vitro* (7, 38), despite the presence of an NPXpY autophosphorylation site at Tyr-960 (39). Consistent with this finding, no tyrosine-phosphorylated IR was bound to the Shc PTB domain *in vitro* (Fig. 2*C*).

It is possible that Shc fails to bind the IR because the NPXpY site is not properly exposed for interaction with the PTB domain. We therefore tested the ability of a Tyr(*P*)-

Table 1. Peptide binding to the Shc PTB domain

| Tyr(*P*)-containing peptide | $IC_{50}$, nM |
|---|---|
| HIIENPQpYFSD | 175 |
| HIIEAPQpYFSD | 80,000 |
| HIIENAQpYFSD | 2,500 |
| HAIENPQpYFSD | 20 |
| HIAENPQpYFSD | 250 |
| HAAENPQpYFSD | 475 |
| HASENPQpYFSD | 15,000 |
| YASSNPEpYLSA | 7,000 |
| YAISNPEpYLSA | 90 |

Binding was measured by competition with a polyoma virus middle tumor antigen phosphopeptide (LSLLSNPTpYSVMRSK). Surface plasmon resonance technology was used to evaluate the ability of phosphopeptides derived from sequences around Tyr-490 in the NGFR (HIIENPQpYFSD) and Tyr-960 in the IR (YASSNPEpYLSA) to bind to the Shc PTB domain. Amino acid substitutions introduced into these peptides are shown in boldface type. Peptide concentrations that inhibited binding of a GST–Shc PTB fusion protein to the middle tumor antigen phosphopeptide, immobilized on a Biacore chip, by 50% ($IC_{50}$) are listed.

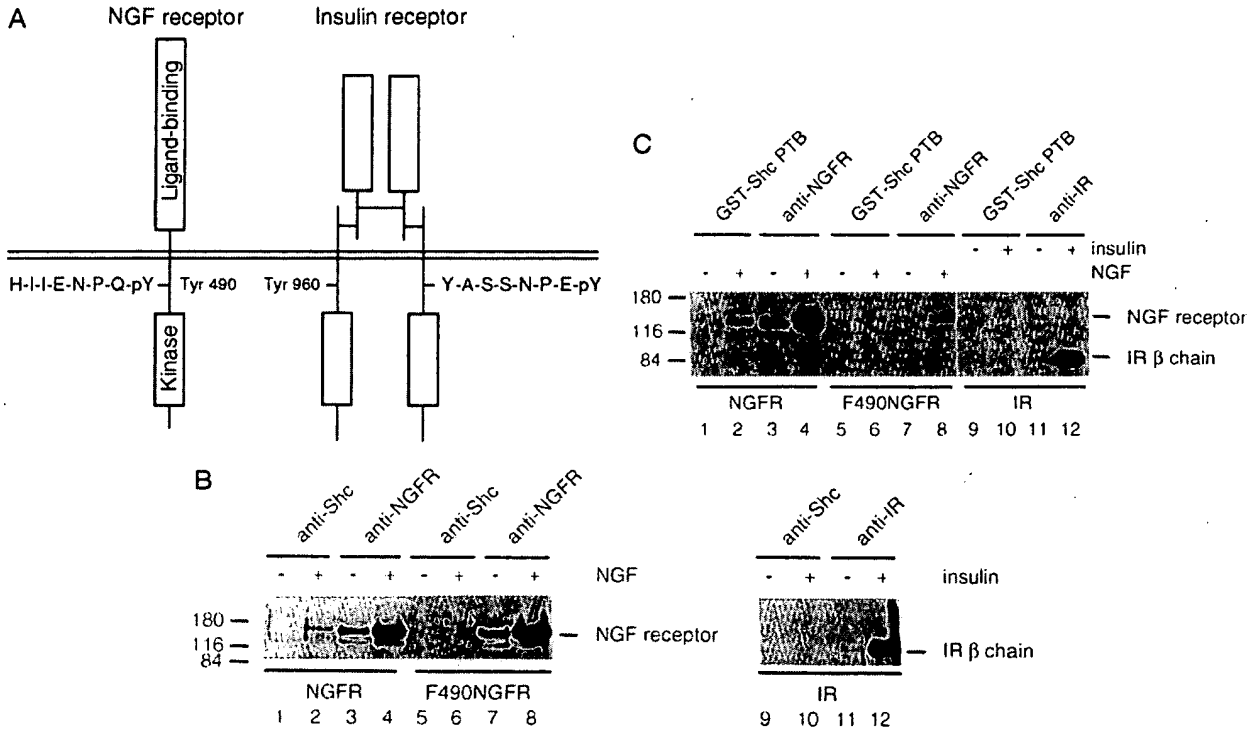



FIG. 2. Shc PTB domain does not stably bind to the NPXpY motif in the IR. (*A*) Both the NGFR and the IR contain an autophosphorylation site within an NPXpY motif in the juxtamembrane domain. In both receptors, the tyrosine residues within these motifs become phosphorylated upon receptor activation, but in contrast to the NGFR, the IR does not stably associate with Shc. (*B*) Anti-Shc immunoprecipitates (lanes 1, 2, 5, 6, 9, and 10) from control (lanes 1, 5, and 9) and growth factor-stimulated (lanes 2, 6, and 10) NIH 3T3 fibroblasts expressing wt (lanes 1 and 2; NGFR) or Phe-490 mutant (lanes 5 and 6; F490NGFR) NGFRs or CHO cells expressing wt IRs (lanes 9 and 10; IR) were examined by anti-Tyr(*P*) immunoblot analysis. Anti-NGFR (lanes 3, 4, 7, and 8) and anti-IR (lanes 11 and 12) immunoprecipitates from control (lanes 3, 7, and 11) and growth factor-stimulated (lanes 4, 8, and 12) cells were analyzed in parallel. (*C*) wt (lanes 1 and 2) and Phe-490 mutant (5 and 6) NGFRs present in lysates from control (lanes 1 and 5) and NGF-stimulated (lanes 2 and 6) cells, and IRs present in lysates from control (lane 9) and insulin-stimulated (lane 10) cells were incubated with GST–Shc PTB fusion proteins bound to glutathione-agarose. Bound proteins were examined by anti-Tyr(*P*) immunoblot analysis. Anti-NGFR immunoprecipitates (lanes 3, 4, 7, and 8) and anti-IR immunoprecipitates (lanes 11 and 12) from control (lanes 3, 7, and 11) and growth factor-stimulated (lanes 4, 8, and 12) cells were analyzed in parallel.

containing peptide based on the sequence around Tyr-960 in the IR to bind the Shc PTB domain (assayed by its ability to compete with the NGFR for binding to the PTB domain). The resulting data show that in contrast to the NGFR phosphopeptide, the IR peptide did not bind to the Shc PTB domain (Fig. 3, lanes 2 and 6, and Table 1). These results indicate that a phosphopeptide based on the IR Tyr-960 autophosphorylation site, which extends 7 residues N-terminal to the Tyr(*P*), has a very low affinity for the Shc PTB domain. Thus the inability of the IR to associate stably with Shc *in vivo* reflects the intrinsically low affinity of the Tyr-960 autophosphorylation site for the Shc PTB domain.

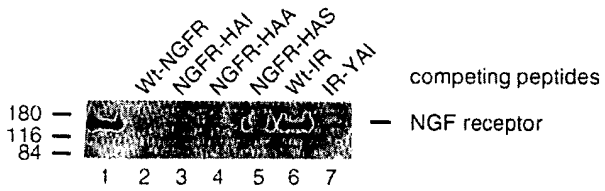



FIG. 3. Aliphatic residue 5 amino acids N-terminal to the Tyr(*P*) is an important determinant for Shc PTB binding. GST–Shc PTB domain fusion proteins bound to glutathione-agarose were incubated with activated NGFRs present in lysates of NGF-stimulated cells in the absence (lane 1) or presence (lanes 2–7) of 2 μM competing wt and mutant Tyr(*P*)-containing peptides based on the sequence around Tyr-490, the Shc PTB binding site in the NGFR (lanes 2–5) or Tyr-960, an autophosphorylation site present within an NPXpY motif in the IR (lanes 6 and 7). Bound proteins were examined by anti-Tyr(*P*) immunoblot analysis. wt NGFR peptide (lane 2), HIIENPQpYFSD; Ala −6 NGFR peptide (NGFR-HAI), HAIENPQpYFSD; Ala −6/Ala −5 NGFR peptide (NGFR-HAA), HAAENPQpYFSD; Ala −6/Ser −5 NGFR peptide (NGFR-HAS), HASENPQpYFSD; wt IR peptide, YASSNPEpYLSA; Ile −5 IR peptide (IR-YAI), YAISNPEpYLSA. Amino acid substitutions introduced in these peptides are shown in boldface type.

A common feature of high-affinity Shc-binding sites with NPXpY motifs is the presence of large aliphatic residues 5 and 6 residues N-terminal to the Tyr(*P*) (positions −5 and −6) (19, 21, 24, 27, 29). The NGFR Tyr-490 site, for example, has Ile at both residues −5 and −6. However, the IR Tyr-960 site has Ser in place of a hydrophobic residue at position −5 (Table 1). To test the possibility that the nature of residues −5 and −6 is important for Shc PTB binding, several substitutions at these positions were made in the NGFR peptide, and these mutant peptides were tested for binding to the Shc PTB domain by using the competition assay described above. Changing the Ile at position −6 to Ala had no inhibitory effect on binding to the PTB domain (Fig. 3). Changing both residues −5 and −6 from Ile to Ala slightly reduced binding to the PTB domain (Fig. 3). Strikingly, changing the Ile residue at position −5 to Ser in addition to changing the Ile at position −6 to Ala, as found at the IR Tyr-960 site, abolished binding of the NGFR phosphopeptide to the Shc PTB domain (Fig. 3). These results were quantitated by using surface plasmon resonance (Table 1). These data implicate the aliphatic residue at position −5 as being important for binding to the Shc PTB domain. This model suggests that the affinity of the IR Tyr-960 autophosphorylation site for the Shc PTB domain might be increased by incorporation of a more hydrophobic residue at position −5. To test this prediction, residue −5 of the IR peptide was

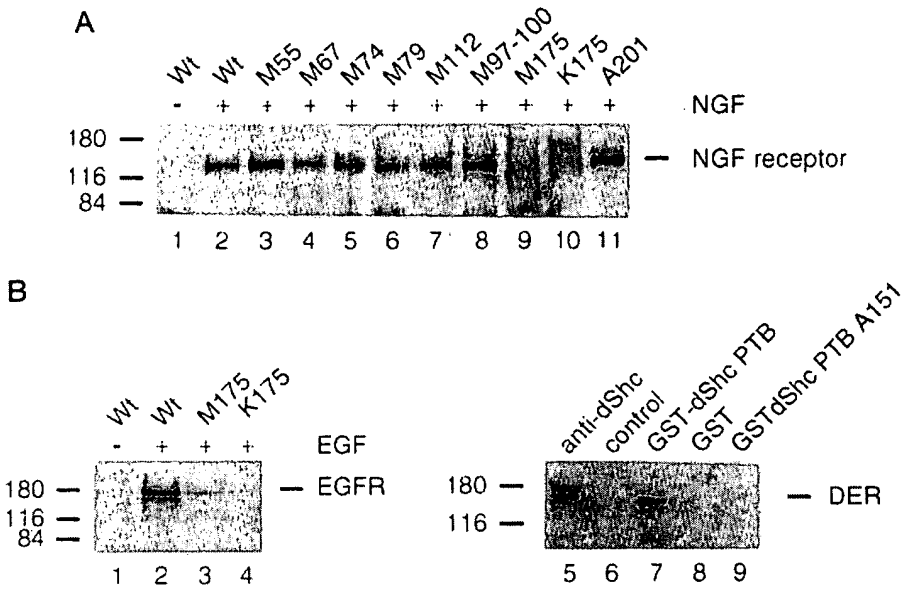



FIG. 4. Requirement for an Arg residue at position 175 in the human Shc PTB domain has been conserved in evolution. (*A*) GST fusion proteins containing wt (lanes 1 and 2) or mutant (lanes 3–11) Shc PTB domains were incubated with NGFRs present in lysates of control (lane 1) and NGF-stimulated (lanes 2–11) cells. Bound proteins were analyzed on an immunoblot with anti-Tyr(*P*). (*B*) Immobilized GST fusion proteins containing wt (lanes 1 and 2) or Met-175 (lane 3) and Lys-175 (lane 4) mutant Shc PTB domains were incubated with lysates from control (lane 1) or EGF-stimulated cells expressing EGFR (lanes 2–4). Immobilized GST (lane 8) and GST fusion proteins containing wt (lane 7) or Ala-151 (lane 9) mutant dShc PTB domain were incubated with fly lysates containing an activated Torso–DER chimeric receptor. Bound proteins were examined by anti-Tyr(*P*) blot analysis. An anti-Shc (lane 5) and a normal rabbit serum immunoprecipitate (lane 6) from the same fly lysates are shown as controls.

changed from Ser to Ile. This mutant IR peptide bound 75-fold more strongly to the PTB domain than its wt counterpart (Fig. 3 and Table 1). These data indicate that Shc PTB binding is modulated by residues outside the NPXpY core binding site.

**Identification of a Conserved Arg Residue in the Shc PTB Domain That Is Required for Binding to Tyr(*P*) Sites.** The identities of residues within the PTB domain that are important for ligand recognition and the mechanism by which the PTB domain binds Tyr(*P*)-containing sites are unknown. The SH2 Tyr(*P*)-binding pocket is composed of basic residues that form hydrogen bonds with the phosphate oxygens or make amino–aromatic interactions with the Tyr ring (40, 41). One invariant Arg of SH2 domains is absolutely essential for Tyr(*P*) binding. (42, 43). To investigate whether basic residues that are conserved in the PTB domains of different Shc family members might also be important for binding Tyr(*P*) sites, all conserved Arg residues in the Shc PTB domain, starting at Arg-55 of $p52^{shc}$ (6), were individually substituted with Met or Ala. Three consecutive Arg residues and the following Lys located at residues 97–100 were changed in combination. GST fusion proteins containing mutant PTB domains were then tested for their ability to bind to the activated NGFR (Fig. 4). Of 10 Arg residues tested, only substitution of Arg-175 had a marked effect on the affinity of the Shc PTB domain for the activated NGFR (Fig. 4*A*). Fusion proteins were expressed at similar levels (results not shown). Substitution of Arg-175 with Met or Lys severely impaired PTB binding to both the NGFR and the EGFR (Fig. 4), indicating that a positive charge in the context of an Arg residue is required at this position. The PTB domain of the *Drosophila* Shc (dShc) protein contains an Arg at residue 151, which is homologous to Arg-175 in human Shc. As was shown (22), the wt dShc PTB domain binds stably *in vitro* to the activated DER, which contains a VDNPEpY site in its C-terminal tail. In contrast, a mutant dShc PTB domain with Arg-151 replaced with Ala was unable to bind efficiently to the activated DER *in vitro* (Fig. 4*B*), suggesting that the requirement for this Arg residue in Tyr(*P*) binding has been conserved in evolution.

## DISCUSSION

**Identification and Characterization of a Core PTB Binding Motif.** The binding of Shc to activated growth factor receptors and cytoplasmic phosphoproteins appears to be an important step in stimulating mitogenic and differentiation pathways. Recent work suggests that the PTB domain, located at the Shc N terminus, plays a significant role in these interactions. Shc frequently associates with proteins that contain NPXpY phosphorylation sites, and phosphopeptides containing the NPXpY sequence can compete for binding of activated growth factor receptors such as the NGFR and EGFR to the Shc PTB domain (19, 21, 25, 27, 29). We have shown (23) that binding of the Shc PTB domain to the NGFR requires autophosphorylation of the receptor at Tyr-490. Here we report that a phosphopeptide modeled on this site binds with high affinity to the Shc PTB domain.

The Tyr-490 NGFR phosphopeptide has been used to analyze the contribution of different residues within the NPXpY motif to PTB binding. The Asn at position −3 appears to be critical for high-affinity binding. The Pro at position −2 is less important for binding *in vitro* although its presence increases the affinity of the interaction approximately 10-fold. This may be important *in vivo*, consistent with the observation that the Pro at position −2 is conserved in several known Shc-binding sites. These findings are in general agreement with those of others that used distinct phosphopeptides or a degenerate phosphopeptide library to analyze the PTB binding site (28, 29, 44).

**PTB Recognition Sites Contain Residues Outside the NPXpY Site That Regulate Binding.** The activated IR, which contains an NPXpY autophosphorylation site, does not bind stably to Shc *in vivo* or *in vitro* (7, 38). Shc, however, is variably phosphorylated in response to insulin, and the Shc PTB domain has been shown to interact with Tyr-960 in the IR by using the two-hybrid method in yeast (44). Shc and the IR may, therefore, engage in a low-affinity interaction but are unable to form a stable complex *in vivo*. These observations suggest that residues outside the NPXpY core PTB binding site affect

receptor–PTB interactions. A comparison of several high-affinity Shc PTB binding sites shows that they possess hydrophobic residues at positions −5 and −6 (Ile in the case of the NGFR), while the IR Tyr-960 site has Ser at position −5. Experiments using NGF and IR phosphopeptides with substitutions at positions −5 and −6 indicate that residue −5 indeed has a critical effect on Shc PTB binding. Ala residues at positions −5 and −6 of the NGFR peptide are tolerated (Table 1). However, the presence of Ser at position −5, as found in the IR, disrupts high-affinity binding of the NGFR peptide to the Shc PTB domain (Table 1). Conversely, the substitution of the Ser at position −5 in the IR phosphopeptide with Ile strikingly increases binding affinity. These data indicate that the Shc PTB domain recognizes the NPXpY sequence and also show that residues N-terminal to the core NPXpY motif are important in determining PTB binding specificity, in a fashion reminiscent of the residues C-terminal to Tyr(*P*) in SH2-binding sites.

**Ligand Recognition by the Shc PTB Domain.** Several Arg residues that are conserved in SH2 domains have been shown to be directly involved in Tyr(*P*) binding (42, 43, 45). We have initiated a characterization of the PTB domain by mutagenesis of Arg residues that are conserved in the PTB domains of different members of the Shc family. This analysis has defined a specific Arg residue near the C terminus of the PTB domain that is important for its interaction with activated growth factor receptors. This Arg is conserved in dShc and is also essential for binding of the dShc PTB domain to activated DER. In SH2 domains, Tyr(*P*) recognition occurs through a direct ion pair between the phosphate moiety and an invariant Arg (45). It is possible that Arg-175 in the Shc PTB domain plays a similar role, in forming part of a Tyr(*P*)-recognition pocket.

Despite the possible similarity in Tyr(*P*) recognition, the general structures of PTB and SH2 domains, and their bound ligands, are likely to be quite different. The low affinity of mutant NGFR phosphopeptides with substitutions of the Asn at position −3 or Pro at position −2 for the Shc PTB domain is consistent with the notion that a reverse-turn secondary conformation is an important structural element for ligand recognition. The magnitude of the effects of the individual Asn and Pro substitutions on binding affinity parallels the known abilities of the relevant amino acids to form a type I β-turn (46). NMR studies of peptides corresponding to the NPXY internalization signal in the low-density lipoprotein receptor and the NPXY autophosphorylation site in the IR have shown that these molecules in solution adopt a type I β-turn with the Asn in position *i* and the Tyr at position *i* + 3 (47). This is quite different from SH2 peptide ligands, which adopt an extended conformation when bound to the SH2 domain (40, 41, 48, 49).

In addition to inferring that the Tyr(*P*) present at position *i* + 3 of a NPXpY type I β-turn characterizes a core motif for PTB binding, our results show that hydrophobic residues N-terminal to this motif play a role in augmenting binding affinity for the Shc PTB domain. These results suggest that the Shc PTB domain has a hydrophobic region that can interact favorably with the Ile at position −5 but is repelled by more polar residues such as Ser.

We thank Drs. I. D. Goldfine and C. Yip for the generous supply of anti-IR monoclonal antibody 51. This work was supported by grants from the Human Frontiers Science Programme and the National Cancer Institute of Canada and from an International Research Scholar Award from the Howard Hughes Medical Institute. P.v.d.G. is a postdoctoral fellow of the National Cancer Institute of Canada. V.K.-M.L. acknowledges a student fellowship from National Sciences and Engineering Research Council. T.P. is a Terry Fox Cancer Research Scientist of the National Cancer Institute of Canada.

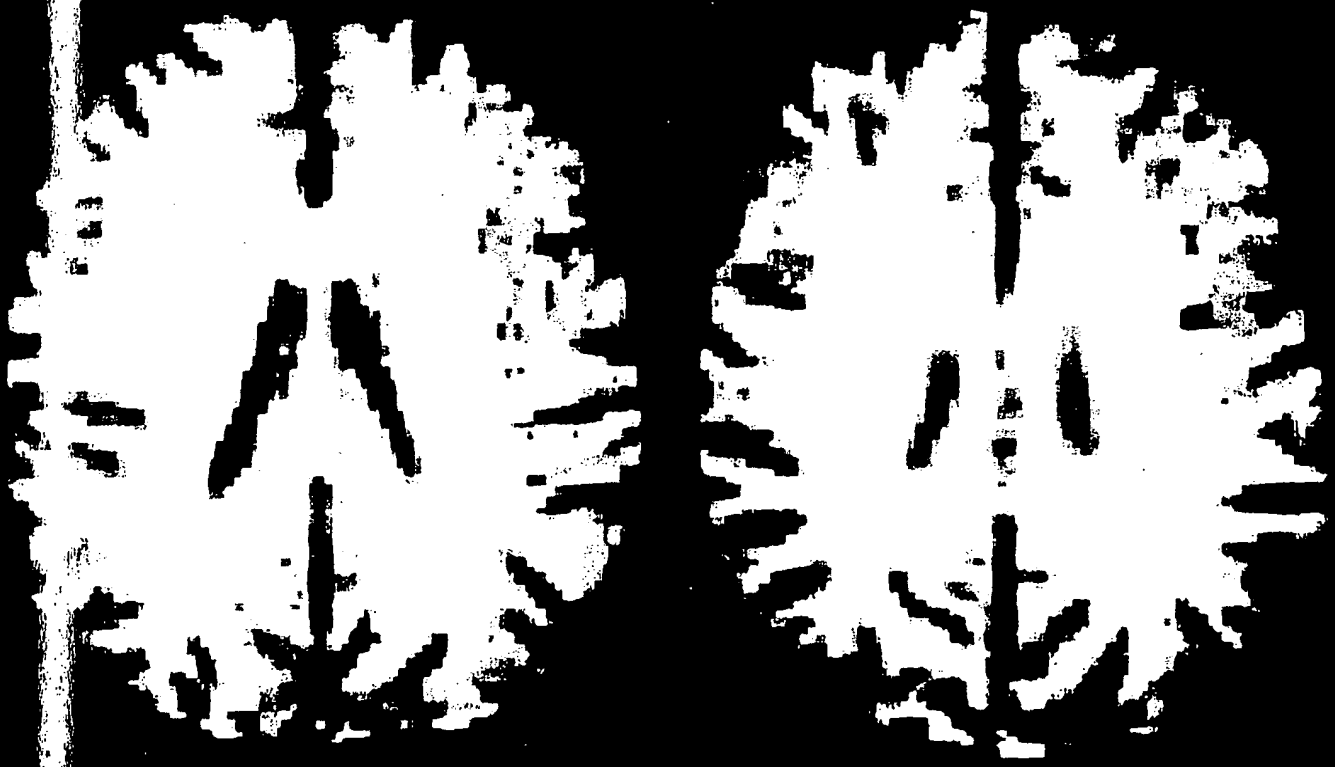
1. Koch, C. A., Anderson, D., Moran, M. F., Ellis, C. & Pawson, T. (1991) *Science* **252**, 668–674.
2. Burns, L. A., Karnitz, L. M., Sutor, S. L. & Abraham, R. T. (1993) *J. Biol. Chem.* **268**, 17659–17661.
3. Crowe, A. J., McGlade, J., Pawson, T. & Hayman, M. J. (1994) *Oncogene* **9**, 537–544.
4. Cutler, R. L., Liu, L., Damen, J. E. & Krystal, G. (1993) *J. Biol. Chem.* **268**, 21463–21465.
5. Lanfrancone, L., Pelicci, G., Brizzi, M. F., Arouica, M. G., Casciari, C., Giuli, S., Pegararo, L., Pawson, T. & Pelicci, P. G. (1995) *Oncogene* **10**, 905–917.
6. Pelicci, G., Lanfrancone, L., Grignani, F., McGlade, J., Cavallo, F., Forni, G., Nicoletti, I., Grignani, F., Pawson, T. & Pelicci, P. G. (1992) *Cell* **70**, 93–104.
7. Pronk, G. J., McGlade, J., Pelicci, G., Pawson, T. & Bos, J. L. (1993) *J. Biol. Chem.* **268**, 5748–5753.
8. Ravichandran, K. S., Lee, K. K., Songyang, Z., Cantley, L. C., Burn, P. & Burakoff, S. J. (1993) *Science* **262**, 902–905.
9. Segatto, O., Pelicci, G., Giuli, S., Digiesi, G., Di, F. P., McGlade, J., Pawson, T. & Pelicci, P. G. (1993) *Oncogene* **8**, 2105–2112.
10. Yokote, K., Mori, S., Hansen, K., McGlade, J., Pawson, T., Heldin, C. H. & Claesson, W. L. (1994) *J. Biol. Chem.* **269**, 15337–15343.
11. Rozakis-Adcock, M., McGlade, J., Mbamalu, G., Pelicci, G., Daly, R., Thomas, S., Brugge, J., Pelicci, P. G., Schlessinger, J. & Pawson, T. (1992) *Nature (London)* **360**, 689–692.
12. Buday, L. & Downward, J. (1993) *Cell* **73**, 611–620.
13. Egan, S. E., Giddings, B. W., Brooks, M. W., Buday, L., Sizeland, A. M. & Weinberg, R. A. (1993) *Nature (London)* **363**, 45–51.
14. Gale, N. W., Kaplan, S., Lowenstein, E. J., Schlessinger, J. & Bar-Sagi, D. (1993) *Nature (London)* **363**, 88–92.
15. Li, N., Batzer, A., Daly, R., Yajnik, V., Skolnik, E., Chardin, P., Bar-Sagi, D., Margolis, B. & Schlessinger, J. (1993) *Nature (London)* **363**, 85–88.
16. Rozakis-Adcock, M., Fernley, R., Wade, J., Pawson, T. & Bowtell, D. (1993) *Nature (London)* **363**, 83–85.
17. Salcini, A. E., McGlade, J., Pelicci, G., Nicoletti, I., Pawson, T. & Pelicci, P. G. (1994) *Oncogene* **9**, 2827–2836.
18. Songyang, Z., Shoelson, S. E., McGlade, J., Olivier, P., Pawson, T., Bustelo, X. R., Barbacid, M., Sabe, H., Hanafusa, H., Yi, T., Ren, R., Baltimore, D., Ratnofsky, S., Feldman, R. A. & Cantley, L. C. (1994) *Mol. Cell. Biol.* **14**, 2777–2785.
19. van der Geer, P., Wiley, S., Lai, V. K.-M., Olivier, J. P., Gish, G., Stephens, R., Kaplan, D., Shoelson, S. & Pawson, T. (1995) *Curr. Biol.* **5**, 404–412.
20. Kavanaugh, W. M. & Williams, L. T. (1994) *Science* **266**, 1862–1865.
21. Blaikie, P., Immanuel, D., Wu, J., Li, N., Yainik, V. & Margolis, B. (1994) *J. Biol. Chem.* **269**, 32031–32034.
22. Lai, K.-M. V., Olivier, J. P., Gish, G. D., Henkemeyer, M., McGlade, J. & Pawson, T. (1995) *Mol. Cell. Biol.* **15**, 4810–4818.
23. Stephens, R. M., Loeb, D. M., Copeland, T. D., Pawson, T., Greene, L. A. & Kaplan, D. R. (1994) *Neuron* **12**, 691–705.
24. Campbell, K. S., Ogris, E., Burke, B., Su, W., Auger, K. R., Druker, B. J., Schaffhausen, B. S., Roberts, T. M. & Pallas, D. C. (1994) *Proc. Natl. Acad. Sci. USA* **91**, 6344–6348.
25. Batzer, A. G., Blaikie, P., Nelson, K., Schlessinger, J. & Margolis, B. (1995) *Mol. Cell. Biol.* **15**, 4403–4409.
26. Obermeier, A., Bradshaw, R. A., Seedorf, K., Choidas, A., Schlessinger, J. & Ullrich, A. (1994) *EMBO J.* **13**, 1585–1590.
27. van der Geer, P. & Pawson, T. (1995) *Trends Biochem. Sci.* **20**, 277–280.
28. Songyang, Z., Margolis, B., Chaudhuri, M., Shoelson, S. E. & Cantley, L. C. (1995) *J. Biol. Chem.* **270**, 14863–14866.
29. Kavanaugh, W. M., Turck, C. W. & Williams, L. T. (1995) *Science* **268**, 1177–1179.
30. Dikic, I., Bater, A. G., Blaikie, P., Obermeier, A., Ullrich, A., Schlessinger, J. & Margolis, B. (1995) *J. Biol. Chem.* **270**, 15125–15129.
31. White, M. F., Livingston, J. N., Backer, J. M., Lauris, V., Dull, T. J., Ullrich, A. & Kahn, C. R. (1988) *Cell* **54**, 641–649.
32. Honegger, A. M., Dull, T. J., Felder, S., Van Obberghen, E., Bellot, F., Szapary, D., Schmidt, A., Ullrich, A. & Schlessinger, J. (1987) *Cell* **51**, 199–209.
33. Forsayeth, J. R., Montemurro, A., Maddux, B. A., DePirro, R. & Goldfine, I. D. (1987) *J. Biol. Chem.* **262**, 4134–4140.
34. Roth, R. A., Cassell, D. J., Wong, K. Y., Maddux, B. A. & Goldfine, I. D. (1982) *Proc. Natl. Acad. Sci. USA* **79**, 7312–7316.

35. Hempstead, B. L., Rabin, S. J., Kaplan, L., Reid, S., Parada, L. & Kaplan, D. R. (1992) *Neuron* **9**, 883–896.
36. Puil, L., Liu, J., Gish, G. D., Mbamalu, G., Bowtell, D., Pelicci, P. G., Arlinghaus, R. & Pawson, T. (1994) *EMBO J.* **13**, 764–773.
37. Dilworth, S. M., Brewster, C. E., Jones, M. D., Lanfrancone, L., Pelicci, G. & Pelicci, P. G. (1994) *Nature (London)* **367**, 87–90.
38. Kovacina, K. S. & Roth, R. A. (1993) *Biochem. Biophys. Res. Commun.* **192**, 1303–1311.
39. Feener, E. P., Backer, J. M., King, G. L., Wilden, P. M., Sun, X. J., Kahn, C. R. & White, M. F. (1993) *J. Biol. Chem.* **268**, 11256–11264.
40. Eck, M. J., Shoelson, S. E. & Harrison, S. C. (1993) *Nature (London)* **362**, 87–91.
41. Waksman, G., Shoelson, S. E., Pant, N., Cowburn, D. & Kuriyan, J. (1993) *Cell* **72**, 779–790.
42. Marengere, L. E. M. & Pawson, T. (1992) *J. Biol. Chem.* **267**, 22779–22786.
43. Mayer, B. J., Jackson, P. K., Van Etten, R. A. & Baltimore, D. (1992) *Mol. Cell. Biol.* **12**, 609–618.
44. Gustafson, T. A., He, W., Craparo, A., Schaub, C. D. & O'Neill, T. J. (1995) *Mol. Cell. Biol.* **15**, 2500–2508.
45. Waksman, G., Kominos, D., Robertson, S. C., Pant, N., Baltimore, D., Birge, R. B., Cowburn, D., Hanafusa, H., Mayer, B. J., Overduin, M., Resh, M. D., Rios, C. B., Silverman, L. & Kuriyan, J. (1992) *Nature (London)* **358**, 646–653.
46. Wilmot, C. M. & Thornton, J. M. (1988) *J. Mol. Biol.* **203**, 221–232.
47. Bansal, A. & Gierasch, L. M. (1991) *Cell* **67**, 1195–1201.
48. Pascal, S. M., Singer, A. U., Gish, G., Yamazaki, T., Shoelson, S. E., Pawson, T., Kay, L. E. & Forman-Kay, J. D. (1994) *Cell* **77**, 461–472.
49. Lee, C.-H., Kominos, D., Jacques, S., Margolis, B., Schlessinger, J., Shoelson, S. E. & Kuriyan, J. (1994) *Structure* **2**, 423–438.

# nature

INTERNATIONAL WEEKLY JOURNAL OF SCIENCE

Volume 373 No. 6515 16 February 1995 $8.50

# Gender and language

**Prospects for safer X-ray imaging**

**Protein motifs and signalling**

**Impact chemistry on Jupiter**

16 Feb 1995 Vol. 373 Issue no. 6515

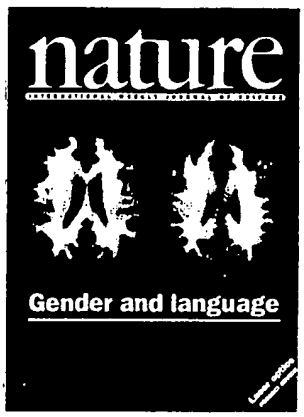
## ON THE COVER

**Gender and language**
Pages 607

**Prospects for safer X-ray imaging**
Page 596

**Protein motifs and signalling**
Page 573

**Impact chemistry on Jupiter**
Page 592




Nature Publishing Company

◀ These composite magnetic resonance images show the distribution of active areas in the brains of males (left) and females during a 'rhyming task'. In males activation is lateralized to the left interior frontal regions but in females the same region is active bilaterally. A long-suspected sex difference in the functional organization of the brain for language is confirmed and shown to occur at the level of phonological processing — see pages 607 and 561.

## THIS WEEK...

### A softer touch with X rays

Phase-contrast techniques greatly increased the flexibility of light microscopy, for example by allowing the observation of thin biological systems *in vivo* and without staining. An analogous technique for high-energy ('hard') X-rays, based on the measurement of phase variations using diffraction from silicon crystals, could now bring phase-contrast imaging to the realms of industrial and medical radiography. Hard X-rays penetrate carbon-based material such as soft tissue with little absorption, hence giving poor contrast. By using phase-contrast to enhance the contrast for weakly absorbing materials it should be possible to obtain clinically useful images at lower — and safer — doses. The example above shows conventional and phase-contrast X-ray views of a mosquito. Page 595.

### Proteins behind signalling networks

Cell–cell interactions are central to the function of multicellular organisms and breakdowns in the communication networks between cells are implicated in cancer, diabetes and many other conditions. But these signals must be relayed from the surface of the cells affected to their intracellular targets. A recent convergence of genetic, biochemical and structural data has focused attention on the conserved protein domains that regulate intracellular signal transduction through their ability to mediate protein–protein interactions. In this week's Review Article, Tony Pawson summarizes current knowledge on signalling networks concentrating on the role of these modular signalling domains. Page 573.

### Shoemaker–Levy 9's impact

Last July's collision of the fragments of comet Shoemaker–Levy 9 with Jupiter dramatically modified the composition and thermal properties of the jovian atmosphere at the impact sites. Shock-induced chemical reactions associated with the impacts of the largest cometary fragments generated large amounts of carbon monoxide, and two compounds not previously detected on Jupiter, carbonyl sulphide and carbon monosulphide. Cooling of the stratosphere for weeks after the impacts may reflect the presence of minor compounds that radiate efficiently at infrared and submillimetre wavelengths. Page 592.

## OPINION



## NEWS



## CORRESPONDENCE



## NEWS AND VIEWS



*Nature*® (ISSN 0028-0836) is published weekly on Thursday, except the last week in December, by Macmillan Magazines Ltd (4 Little Essex Street, London WC2R 3LF). Annual subscription for USA and Canada US$425 (institutional/corporate), US$135 (individual making personal payment). US and Canadian orders to: *Nature*, Subscription Dept, PO Box 1733, Riverton, NJ 08077-7333, USA. Other orders to *Nature*, Brunel Road, Basingstoke, Hants RG21 2XS, UK. Second class postage paid at New York, NY 10012, and additional mailing offices. Authorization to photocopy material for internal or personal use, or internal or personal use of specific clients, is granted by *Nature* to libraries and others registered with the Copyright Clearance Centre (CCC) Transactional Reporting Service, provided the base fee of $9.00 an article (or $1.00 a page) is paid direct to CCC, 27 Congress Street, Salem, MA 01970, USA. Identification code for *Nature*: 0028-0836/95 $9.00+$1.00. US Postmaster send address changes to: Nature, 65 Bleecker Street, New York, NY 10012. Published in Japan by Nature Japan K.K., Shin-Mitsuke Bldg. 36 Ichigaya Tamachi, Shinjuku-ku, Tokyo 162, Japan. 

# Protein modules and signalling networks

Tony Pawson

**Communication between cells assumes particular importance in multicellular organisms. The growth, migration and differentiation of cells in the embryo, and their organization into specific tissues, depend on signals transmitted from one cell to another. In the adult, cell signalling orchestrates normal cellular behaviour and responses to wounding and infection. The consequences of breakdowns in this signalling underlie cancer, diabetes and disorders of the immune and cardiovascular systems. Conserved protein domains that act as key regulatory participants in many of these different signalling pathways are highlighted.**

CELLULAR interactions can be viewed as proceeding in two steps. Initially, an extracellular molecule binds to a specific receptor on a target cell, converting the dormant receptor to an active state. Subsequently, the receptor stimulates intracellular biochemical pathways leading to a cellular response, which may involve progression through the cell cycle and changes in cellular gene expression, cytoskeletal architecture, protein trafficking, adhesion, migration and metabolism. How are these internal pathways controlled and organized? Recently, a convergence of genetic, biochemical and structural data has focused attention on conserved protein modules that regulate signal transduction through their ability to mediate protein-protein interactions. These conserved modules represent common regulatory features of many distinct signalling pathways which are used to build up complex networks of interacting proteins.

## SH2, SH3 and PH domains as building blocks

Many polypeptide hormones, as well as cytokines, antigens and components of the extracellular matrix, bind membrane-spanning receptors that signal through associated cytoplasmic protein kinase (PTK) domains. Although the targets of these PTKs may have quite different biochemical activities and biological functions, they often contain related sequences of 50–100 amino acids in length. These sequences, referred to as Src-homology-2 (SH2), Src-homology-3 (SH3) and pleckstrin homology (PH) domains[1–3], can each fold into a compact and functional module independently of surrounding sequences. SH2 and SH3 domains each recognize short peptide motifs bearing phosphotyrosine (pTyr) in the case of SH2, or one or more proline residues in the case of SH3. These conserved features are embedded in more variable sequences responsible for the high affinity and specificity with which individual SH2 or SH3 domains bind (Fig. 1). The binding properties of PH domains are less well understood, although they may associate with phospholipids or bind specific proteins. They may well promote the association of signalling proteins with membranes.

As they recognize pTyr, SH2 domains are explicitly involved in tyrosine kinase signalling pathways. Various combinations of SH3 and PH domains are frequently found in the same polypeptides as SH2 domains (Fig. 2), and can collaborate with them in forming signalling complexes downstream of PTKs. Indeed, proteins such as Grb2 and Crk are composed almost entirely of SH2 and SH3 domains, and apparently function as molecular adaptors to nucleate the formation of protein complexes, as discussed below. Polypeptides with multiple SH2, SH3 and PH domains can clearly engage in complex protein-protein interactions (Fig. 1*a*), which can regulate many facets of the signalling process, including enzyme-substrate recognition, enzymatic activity, and subcellular localization. The essential biochemical and structural properties of SH2, SH3 and PH domains are now described.

**SH2 and PTB domains.** Proteins with SH2 domains control biochemical pathways involving phospholipid metabolism, tyrosine phosphorylation and dephosphorylation, activation of Ras-like GTPases, gene expression, protein trafficking and cytoskeletal architecture (Fig. 2)[4]. *In vivo*, SH2-containing proteins bind pTyr-containing sites on activated receptors and cytoplasmic phosphoproteins[5–7]. These interactions can be mimicked *in vitro* using isolated SH2 domains and short phosphopeptides of 5–10 amino acids (Fig. 1*b*). SH2 domains bind to phosphopeptides of optimal sequence with relatively high affinity ($K_d$ = 10–100 nM), and to phosphopeptides of random sequence with ~1,000-fold lower affinity[8–10]. They have essentially no affinity for unphosphorylated peptides. Most of the binding energy thus derives from pTyr recognition, although the amino acids surrounding the pTyr can increase the affinity by three orders of magnitude[8]. The phosphopeptide binding site is bipartite[11–14]. A conserved pocket lined by basic residues binds the pTyr; this pocket contains the only invariant SH2 residue, an arginine which forms hydrogen bonds with two pTyr phosphate oxygens. The second binding surface is more variable, and allows specific recognition of the amino acids immediately C-terminal to the pTyr[15]. One group of SH2 domains, typified by those of Src and Lck, make specific contacts with the three residues immediately following the pTyr (the +1 to +3 residues; Fig. 1*b*). Such SH2 domains prefer hydrophilic amino acids at the +1 and +2 positions, but have a small hydrophobic pocket which accommodates a hydrophobic residue at the +3 position. The Src SH2 domain thus selects the sequence pTyr-Glu-Glu-Ile from a degenerate phosphopeptide library. By contrast, the SH2 domains of proteins such as the phospholipase PLC-$\gamma$1 and the Syp/SH-PTP2 protein-tyrosine phosphatase (PTPase) recognize at least five, primarily hydrophobic residues following the pTyr, which fit into an extended hydrophobic groove running across the ligand-binding surface.

The bipartite organizations of the SH2 ligand-binding site allows tyrosine phosphorylation to function as an all-or-none switch for SH2 binding, while enabling the sequence context of the pTyr site to dictate which SH2 domains (and therefore which SH2 signalling proteins) are bound. Binding to pTyr sites can affect SH2-containing proteins in multiple ways, including direct stimulation of enzymatic activity[16], relocalization within the cell[17], and enhanced tyrosine phosphorylation[18].

Recent evidence indicates that Shc and the insulin-receptor substrate IRS-1 may contain a novel pTyr-binding (PTB) domain of ~200 residues which bears only a very tentative relationship to SH2 sequences (refs 19, 20, 129; P. van der Geer *et al.*, manuscript submitted). The PTB domain appears to recognize motifs with the consensus Asn-Pro-X-pTyr, and to require residues N-terminal to pTyr for binding. Although the physiological significance of this domain is not yet known, it appears important for the interactions of Shc and IRS-1 with activated receptors (ref. 129; P. van der Geer *et al.*, submitted).

**SH3 domains.** SH3 domains are found in many proteins involved in tyrosine kinase signalling, but also in cytoskeletal components and subunits of the neutrophil cytochrome oxidase,

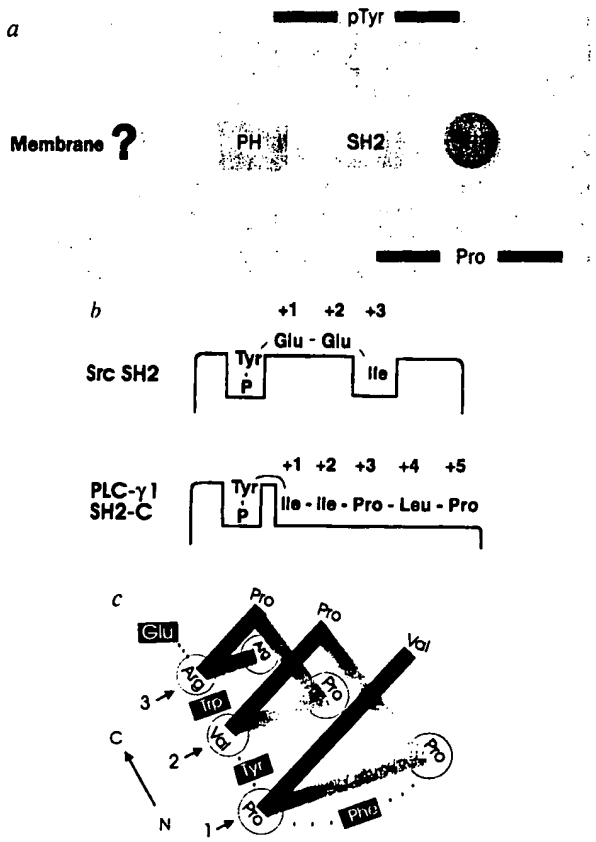



FIG. 1 SH2, SH3 and PH modules can form networks of interacting proteins. *a*, A protein containing SH2, SH3 and PH domains can potentially form multiple protein complexes. *b*, SH2 domains recognize pTyr-containing sites. Phosphorylation regulates binding, whereas the residues C-terminal to the tyrosine impart specificity. Peptides with the sequence pTyr-Glu-Glu-Ile bind with high affinity to the Src SH2 domain, which has a basic pTyr-binding pocket and a hydrophobic pocket for the +3Ile. The peptide pTyr-Ile-Ile-Pro-Leu-Pro-Asp binds the C-terminal PLC-γ1 SH2 domain, which possesses an extended hydrophobic groove for the C-terminal residues. Structures for both these complexes have been solved[11,13], and are shown here in a highly schematic form. *c*, SH3 domains recognize proline-rich motifs. A proline-rich peptide derived from Sos is shown adopting a PPII helix as it binds the surface of a Sem-5/Drk/Grb2 SH3 domain in a class II orientation. Peptide residues that contact the SH3 domain are circled. Some of the SH3 residues that form the ligand-binding site are boxed. See text and refs 26–29 for details.

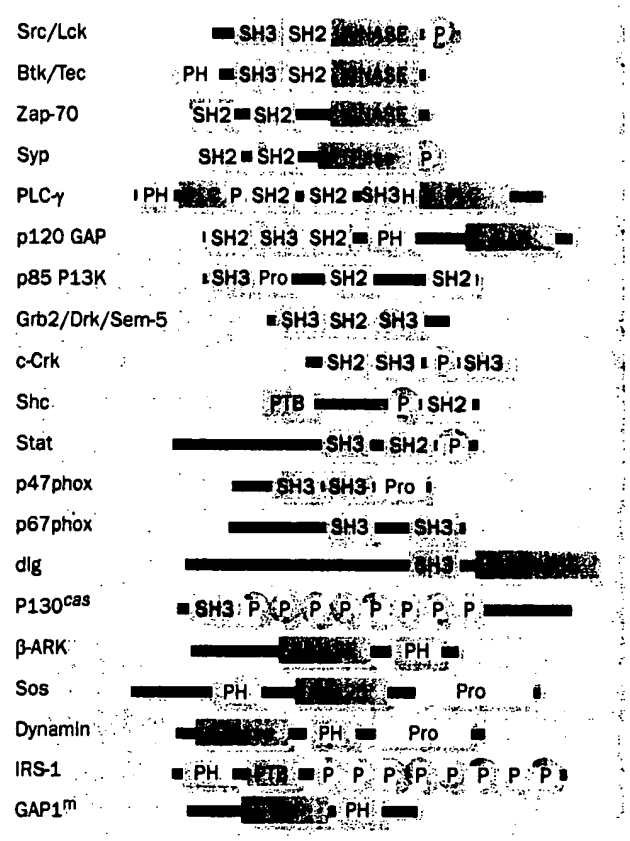



FIG. 2 Selected proteins (see text) with SH2, PTB, SH3 and PH domains. Pro. proline-rich SH3-binding sites; P, phosphotyrosine-containing SH2-binding site; PTPase, phosphotyrosine phosphatase domain; G.KINASE, guanylate-kinase-related domain; Cdc25, Ras guanine-nucleotide-exchange domain.

among others (Fig. 2)[21,22]. SH3-binding sites consist of proline-rich peptides of approximately 10 amino acids (Fig. 1*c*)[23,24], which bind to isolated SH3 domains with dissociation constants of 5–100 μM (ref. 25). Recent structural and mutagenic analysis of peptide-SH3 complexes[26-30] shows that peptides associated with SH3 domains adopt a left-handed polyproline type II helix, with three residues per turn. The peptide ligand therefore has three spines, two contacting the SH3 domain, and the third stabilizing the PPII helix. The core ligand appears to be a seven-residue peptide containing the consensus X-P-p-X-P, where X tends to be an aliphatic residue and the two conserved prolines (P) are crucial for high affinity binding. The intervening scaffolding residue (p) also tends to be a proline. Each X-P pair fits into a hydrophobic pocket formed by conserved SH3 aromatic residues (sites 1 and 2), providing the principal binding energy. A third pocket (site 3) is more variable, although it frequently binds an arginine. In the peptide NH2-*R*-A-L-P-P-L-P-R-Y-COOH, for example, which binds the Src SH3 domain, the first (underlined) arginine forms a salt bridge with an aspartate in the SH3 (ref. 26). The corresponding residue in the Abl SH3 domain is a threonine, which apparently dictates the selection of more hydrophobic residues such as methionine or tryptophan at this site in the target sequence[28,30].

SH3-binding peptides are pseudo-symmetrical and, remarkably, can potentially bind in either orientation (that is, amino- to carboxy-terminal as above (class I), or carboxy- to amino-terminal (class II)[26,27,29]. In the case of a class II ligand for the Src SH3 domain, the arginine fitting into the site-3 pocket lies at the C-terminal end of the peptide, rather than at its N terminus. Hence the peptide $NH_2$-A-F-A-P-P-L-P-R-*R*-COOH binds with reasonable affinity to the Src SH3 domain, but in a class II orientation[26]. The Ras guanine-nucleotide-exchange factors mSos1 and mSos2 are striking examples of physiological class II ligands. Sos peptides bound to an SH3 domain of Grb2, or its invertebrate homologue Sem-5, are complexed in the C- to N-terminal orientation, with an arginine forming a salt bridge to a glutamate in site 3 of the SH3 domain[27,29]. In Sos SH3-binding sites the C-terminal arginine is usually followed by a further run of arginine residues, increasing the affinity for the Grb2 SH3 domains[26]. The capacity of SH3 domains to bind ligands in either orientation both expands the range of potential binding partners, and has fascinating biological implications. For example, the ligand orientation will determine the spatial organization of the resulting complex, which may be critical for signalling.

Studies using degenerate peptide libraries and phage display libraries have indicated that each SH3 domain has a distinct binding preference[24,25,30]. Specificity is apparently conferred by the interactions between non-proline residues in the ligand and two variable SH3 loops flanking the main hydrophobic binding surface.

There is no evidence that SH3-binding depends on a modification like phosphorylation, but the conformation of the SH3-containing protein or its binding partner may control the accessibility of the SH3 domain or its binding site. Potential binding partners may also be spatially separated, only interacting when juxtaposed by relocation. In addition, several SH3-binding sites have consensus sites for phosphorylation by proline-directed kinases, such as MAP kinase, which may affect the contacts made with the SH3 domain[31].

**PH domains—new kids on the block.** PH domains have been found in a wide range of proteins, including serine/threonine- and tyrosine-specific protein kinases, substrates for these kinases, phospholipase C isoforms, regulators of small GTPases, the GTPase dynamin, and cytoskeletal proteins[32-34] (Fig. 2). They appear to have a common structure, comprised of two roughly perpendicular anti-parallel $\beta$-sheets, followed by a C-terminal amphipathic $\alpha$-helix[35]. The loops connecting the $\beta$-strands are highly variable in length and sequence, and lie at the opposite face of the domain from the $\alpha$-helix. The nature of PH domain ligands has been contentious. Phospholipids, notably phosphatidylinositol-4,5-$P_2$, have been reported to associate with the positively charged 'variable loop' face of the PH domain[36]. X-ray crystallographic structures of the dynamin PH domain, however, reveal no extended hydrophobic cavity that could accommodate lipid side chains[37,38]. PH domains have also been reported to bind the $\beta/\gamma$ subunits of heterotrimeric G proteins through their $\alpha$-helix and extended C-terminal sequences[39], as well as protein kinase C[40]. The true nature of their physiological ligands, be they phospholipids, proteins, or both, remains to be established.

Several lines of circumstantial evidence suggest that PH domains may tether signalling proteins to membranes, a property consistent with the candidate ligands discussed above. First, PH-containing proteins are generally cytoplasmic, but associate with the plasma membrane. Indeed, cytoplasmic PTKs with N-terminal PH domains like Btk, lack the myristylated N terminus that anchors Src and its relatives to the membrane (Fig. 2). Similarly, the C-terminal region of $\beta$-adrenergic receptor kinase, which includes a PH domain, is crucial for recruiting the kinase to its membrane-bound substrate[41]. In the closely related rhodopsin kinase, the PH domain is replaced by a C-terminal farnesylation site. The main substrate for the insulin receptor, IRS-1, which acts as an intermediate SH2-docking protein in insulin signalling, also contains a PH domain which may play a role in juxtaposing the kinase and its substrate at the membrane. Although the PH domain is not essential for mitogenic signalling on insulin stimulation, it enhances IRS-1 phosphorylation in cells expressing low levels of the receptor, a property consistent with membrane localization (M. G. Myers and M. White, personal communication). The best evidence for PH domain function, however, comes from a genetic source. A missense mutation in the X-linked *btk* gene that alters an amino acid in the Btk PH domain impairs murine B-cell development[42].

## Receptors coupled through pTyr–SH2

**Receptor tyrosine kinases.** In many receptors, the extracellular ligand-binding domain and the cytoplasmic tyrosine kinase domain are part of the one polypeptide. Binding of the appropriate growth factor is accompanied by receptor dimerization and autophosphorylation[43]. In the receptors for epidermal growth factor and platelet-derived growth factor (EGFR and PDGFR, respectively), autophosphorylation sites generally lie in non-catalytic regions of the cytoplasmic domain, between the membrane and kinase domain, in a non-catalytic sequence within the kinase domain, or in the C-terminal tail. A principal function of these sites is to bind the SH2 domains of specific receptor targets[5], stimulating their activities. The $\beta$-PDGFR has multiple autophosphorylation sites, each of which is relatively specific for a particular SH2 protein[7] (Fig. 3*a*). Mutant receptors lacking one or more autophosphorylation sites can therefore be used to examine the effects of uncoupling the receptor from specific signalling pathways. Phosphorylation of Tyr 1,021 in the receptor tail, for example, induces binding of PLC-$\gamma$1; however a mutant receptor in which Tyr 1,021 is replaced by Phe cannot bind PLC-$\gamma$1 or stimulate hydrolysis of phosphatidylinositol-4,5-$P_2$, although it interacts normally with other SH2 signalling proteins[44-46]. These and similar experiments have suggested that multiple receptor targets, including phosphatidylinositol-3-OH kinase (PI3K), Src and PLC-$\gamma$1, contribute to efficient mitogenic signalling[47,48]. Some of the affected targets may be functionally redundant, especially at the level of Ras activation[49].

Several receptor PTKs have similar features, although there are variations on the theme, and not all receptor SH2-binding sites show the striking specificity of the $\beta$-PDGFR. The receptor for hepatocyte growth factor, Met, has two closely spaced autophosphorylation sites, each of which binds multiple SH2-containing proteins[50]. Substitution of these two sites abolishes Met mitogenic activity. By contrast, the main insulin-receptor substrate, IRS-1, has PH and PTB domains but is devoid of SH2 sequences[51]. It does, however, contain at least 20 potential tyrosine phosphorylation sites which, when occupied, form complexes with the SH2 domains of signalling proteins such as PI3K, Grb2, Syp and Nck[52], and may therefore act as an intermediate SH2-docking protein, with more potential SH2-binding sites than can be accommodated within the receptor itself.

**Multisubunit receptors.** Antigen receptors in lymphoid cells employ similar interactions and signalling pathways, but with the ligand-binding chains, the tyrosine kinase catalytic activity, and the tyrosine-phosphorylated SH2-docking sites on separate polypeptides (Fig. 3*b*). T-cell activation involves a series of interactions between the antigen-presenting cell and the T cell. The polymorphic $\alpha/\beta$ chains of the T-cell antigen receptor (TCR) recognize specific antigenic peptides associated with major histocompatibility complex (MHC) proteins on the surface of the antigen-presenting cell. The MHC molecule associates with CD4 or CD8 proteins spanning the T-cell membrane, which bind the Src-family kinase, Lck, in the T-cell cytoplasm. The TCR itself forms non-covalent interactions with a series of transmembrane signalling subunits ($\gamma$, $\delta$, $\varepsilon$ and $\zeta$), each of which contains one to three copies of the motif Tyr-X-X-Leu/Ile-$X_{6-8}$-Tyr-X-X-L/Ile[53]. Phosphorylation of the tyrosines within these motifs in antigen-stimulated T cells (probably by Lck[54]) creates binding sites for the SH2 domains of downstream-signalling proteins, notably another tyrosine kinase termed Zap-70[54]. Zap-70 has two adjacent SH2 domains at its N terminus that may engage the closely spaced pTyr residues in each motif[55], activating the Zap-70 kinase domain. This in turn stimulates PLC-$\gamma$1, leading to a rise in intracellular calcium.

The central role of Zap-70 in this pathway has been confirmed by a rare human immunodeficiency, in which $CD8^+$ cells are not produced, and the remaining $CD4^+$ peripheral T-cells do not produce IL-2 in response to antigenic stimulation. Without bone marrow transplantation, patients succumb to opportunistic infections at an early age. All patients studied so far have mutations in both *zap-70* alleles, disrupting the coding sequence for the kinase domain and preventing the production of detectable Zap-70 protein[56,57].

Although stimulation of this TCR signalling pathway is essential for T-cell activation, it cannot alone induce IL-2 production in quiescent T-lymphocytes. The necessary second signal can be delivered by the association of B-7, an immunoglobulin-like molecule on the surface of the antigen-presenting cell, and CD28, a transmembrane protein on T cells. CD28 has a short cytoplasmic tail containing the motif Tyr-Met-Asn-Met. Recent evidence indicates that the association of B-7 and CD28 induces the phosphorylation of CD28 at this site, allowing it to associate with the SH2 domain(s) of PI3K[58,59]. Substitution of this Tyr with Phe abolishes CD28's ability to synergize with the T-cell receptor in stimulating IL-2 expression, indicating that its association with PI3K or other SH2 proteins is biologically important.

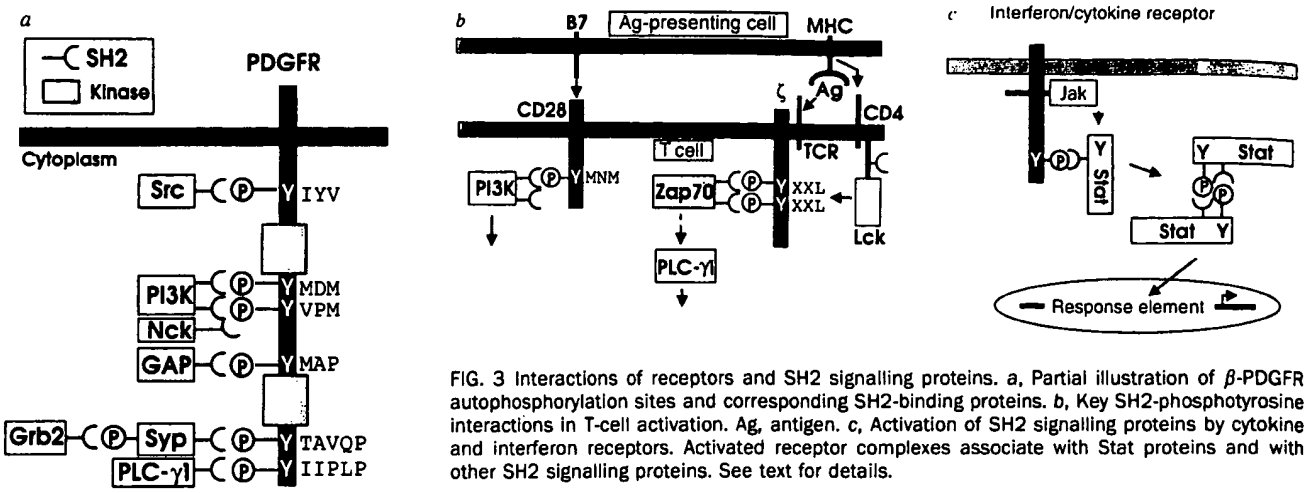



FIG. 3 Interactions of receptors and SH2 signalling proteins. *a*, Partial illustration of β-PDGFR autophosphorylation sites and corresponding SH2-binding proteins. *b*, Key SH2-phosphotyrosine interactions in T-cell activation. Ag, antigen. *c*, Activation of SH2 signalling proteins by cytokine and interferon receptors. Activated receptor complexes associate with Stat proteins and with other SH2 signalling proteins. See text for details.

In the absence of CD28, TCR stimulation induces a non-responsive state termed anergy, rather than activation.

In T cells, the activation of SH2 proteins therefore involves multiple subunits, and requires several discrete contacts with the antigen-presenting cell, each activating a distinct subset of SH2-regulated signalling pathways. This may represent a fail-safe mechanism for preventing the inappropriate activation of T-cells.

**Cytokine receptors.** The receptors for cytokines and interferons are composed of one to three polypeptides and are constitutively associated with members of the Janus protein kinase (Jak) family of intracellular PTKs (although the amount of associated Jak may increase on ligand binding[60,64]).

Ligand-binding induces receptor oligomerization and concomitant Jak activation, possibly by cross-phosphorylation. The resulting tyrosine phosphorylation of the receptor signalling subunits enables them to bind SH2-containing proteins such as PLC-γ1, PI3K, Shc, Syp and PTP-1C, which are in turn phosphorylated by the receptor complex[65,66]. The PTK substrates most directly implicated in cytokine signalling, however, are transcription factors termed Stat proteins[67,68]. These polypeptides possess an SH2 domain, through which they bind specific pTyr sites on the activated receptor. Each subfamily of cytokine receptors regulates a unique set of Stat proteins in a fashion that is determined by the SH2-binding specificity of the receptor phosphorylation sites[130]. Subsequent phosphorylation of the Stat proteins at a tyrosine immediately C-terminal to the SH2 domain[69] induces their dimerization through mutual pTyr-SH2 interactions[70] (Fig. 3*c*). As a consequence, they translocate to the nucleus, bind to specific promoter elements and induce gene expression[60]. Stat phosphorylation, translocation, and DNA-binding activity are all dependent on the invariant SH2 arginine residue. The location of the Stat pTyr site precludes an interaction with its own SH2 domain of the sort proposed for Src family kinases and Crk (see below), but permits dimerization by recognition of a second Stat molecule.

## Pathways and networks

**The Ras pathway.** If proteins with SH2 and SH3 domains indeed mediate intracellular signalling by PTKs, then mutations affecting the ability of these domains to bind pTyr or proline-rich sites should block cellular responses to external cues. This hypothesis is supported by studies of Sem-5/Drk/Grb2, an adaptor protein that is involved in linking tyrosine kinases to Ras. Sem-5 contains an SH2 domain flanked by two SH3 domains, and couples the Let-23 receptor PTK to Ras in the differentiation of vulval progenitor cells in *Caenorhabditis elegans*[71]. Mutations affecting residues in the SH2 or either SH3 domain can uncouple Let-23 from the Ras pathway to a greater or lesser extent. In *Drosophila*, the homologue of Sem-5, Drk, is required for signalling from multiple receptor PTKs, including the *Drosophila* EGFR homologue, and the torso and sevenless receptors[72,73] (Fig. 4*a*). Flies homozygous for loss-of-function mutations in *drk* which alter basic residues forming part of the SH2 pTyr-binding pocket, die as pupae. Unlike wild-type Drk, these mutant Drk proteins neither bind to activated receptors nor become associated with the plasma membrane. Genetic and biochemical data indicate that Drk may act through Sos, a Ras guanine-nucleotide-release protein (GNRP) containing proline-rich motifs within its C-terminal tail, which bind the Drk SH3 domains. Mutations which inhibit binding of the N-terminal Drk SH3 domain to Sos block receptor signalling *in vivo* (T. Raabe *et al.*, submitted). The binding of SH2 and SH3 domains to pTyr sites and proline-rich motifs, respectively, is thus indeed critical for cellular responses mediated by receptor PTKs.

The mammalian homologue of Sem-5 and Drk is Grb2, whose SH3 domains bind mSos1 and mSos2, two mammalian homologues of *Drosophila* Sos[74-78]. Stimulation with growth factors such as EGF induces the Grb SH2 domain to bind motifs with the consensus sequence pTyr-X-Asn-X. Such a site lies at Tyr 1068 of the EGFR[15,79], recruiting Grb2 and Sos into a heterotrimeric complex with the EGFR and potentially stimulating Sos proteins to activate Ras. In metazoans, GTP-bound Ras activates a protein-serine/threonine kinase cascade, eventually triggering phosphorylation of transcription factors by MAP kinase[82].

**Many routes to Ras?** This basic scheme, often depicted as a linear pathway from activated receptors to Ras (Fig. 4*a*), masks a number of complexities. Although the PDGF- and EGF-receptors can bind Grb2 directly[79,83], they also bind and phosphorylate SH2-containing proteins such as Shc and the Syp/SH-PTP2 tyrosine phosphatase, which contain Tyr-X-Asn-X motifs. *In vivo*, the Grb2 SH2 domain binds both to activated receptors, and to phosphorylated Shc and Syp[84,85], allowing Grb2 to form multiple complexes with other SH2-containing proteins as well as activated receptors. Biochemical and genetic data indicate that both the Shc–Grb2 and Syp–Grb2 complexes may stimulate the Ras pathway[7,84,86]. Taken together, these observations point to a complex network of interactions synchronizing Ras activation with other signalling events (Fig. 4*b*). The network may also potentiate activation of the Ras pathway, functioning as a quantitative monitor of receptor activation. In addition, it enables PTKs lacking intrinsic Grb2-binding sites to activate the Ras pathway.

Like its SH2 domain, the SH3 domains of Grb2 can interact with several proteins *in vivo*. As well as mSos1 and 2, these include dynamin, a GTPase that may link Ras signalling to clathrin-mediated endocytosis[87-89]. To add further complexity, mammalian cells contain several Ras GNRPs apart from mSos1

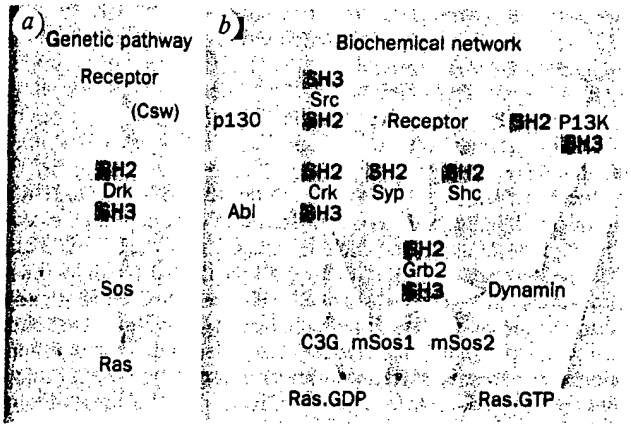



FIG. 4 Signalling through the Ras pathway. A linear pathway or more complex network? *a*, Genetic and biochemical experiments in *Drosophila* suggest a linear pathway, in which receptor PTKs couple to Drk (the homologue of Grb2), possibly with the SH2-containing PTPase Corkscrew (Csw; the homologue of mammalian Syp) acting as an intermediate[125]. Drk then activates Sos. *b*, In mammalian cells, a complex network of interactions has been identified, involving proteins that can regulate Ras, only part of which is shown here. Interactions are as discussed in the text. See Fig. 2 for details of protein structures. The C-terminal proline-rich region of dynamin can bind not only Grb2 and PI3K, but also PLC-γ1. The Crk SH3 domain interacts *in vivo* with both the Ras GNRP C3G, and also with a proline-rich motif in the C-terminal tail of the Abl PTK[126,127]. PI3K makes multiple interactions, with receptor autophosphorylation sites through its SH2 domain, with the SH3 domain of Src family kinases through a proline-rich region of p85, and with GTP-bound Ras[128]. These complexes, along with p85 tyrosine phosphorylation, may act synergistically to regulate PI3K activity.

and mSos2. One of these, C3G, has a C-terminal domain related to other Ras GNRPs, and can activate Ras in yeast[90]. Its central proline-rich sequence binds the N-terminal SH3 domain of the adaptor protein Crk[90,91], and Crk–C3G complexes may thus, like Grb2–Sos complexes, couple tyrosine phosphorylation to Ras activation. Crk can also bind mSos1 *in vivo*[91]. Crk was first identified as a retroviral oncogene[2], and its transforming ability may be partly due to its interactions with C3G. The binding specificity of the Crk SH2 domain differs from that of Grb2 (pTyr-Asp-His-Pro), and may therefore expand the repertoire of tyrosine phosphorylation sites that can activate the Ras pathway (Fig. 4*b*). Engaging both the Crk and Grb2 SH2 domains may even enhance the extent of Ras activation.

Finally, the mechanism of Sos activation may be complex, since changes in its catalytic activity have been hard to detect. One possibility is that Drk/Grb2 simply positions Sos in the membrane next to Ras, and it is clear that Sos membrane association is critical[80,81]. However, recent genetic data in *Drosophila* have raised the possibility of a Drk-independent pathway to Sos (V. Banerjee, personal communication; T. Raabe *et al.*, submitted). Furthermore the N-terminal Sos PH domain seems important for activation, perhaps through interactions with membrane components (M. Czech, personal communication), while the C-terminal Sos tail has an inhibitory function (ref. 81; S. Egan, personal communication). Drk/Grb2 may therefore activate Sos, at least in part, by relieving this repression.

**Interconnecting pathways.** Just as the exchange of GDP for GTP on Ras is controlled by multiple exchange factors, Ras inactivation is also regulated. Three mammalian proteins have been found to inactivate Ras (p120 GTPase-activating protein (GAP), neurofibromin and GAP1$^m$, collectively termed rasGAPs[92,93]) by greatly enhancing the rate at which the GTP bound to activated Ras is hydrolysed to GDP. Enhancing or suppressing the activities of these rasGAPs can therefore affect Ras activation. Mutations in the gene encoding neurofibromin, for example, can in some cases elevate levels of Ras·GTP, thereby inducing aberrant proliferation. In a more physiological setting, transient inhibition of rasGAP activity may increase the amount of Ras·GTP in activated T-cells[94].

Two GAP proteins are directly implicated in the control of Ras by tyrosine kinases. GAP1$^m$ has a PH domain, and is struc-

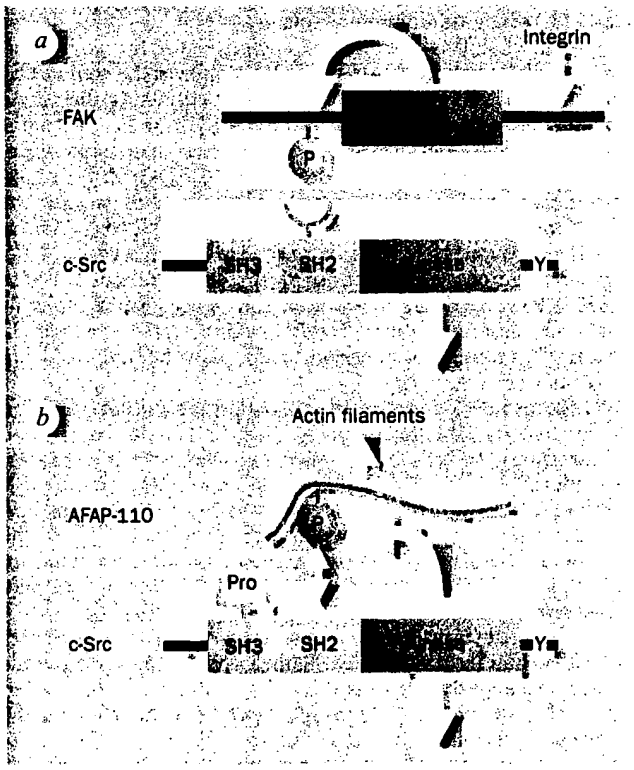



FIG. 5 Src family SH2 and SH3 domains have multiple functions. *a*, The Src SH2 domain binds other PTKs. In this scheme, tyrosine phosphorylation of Fak induces binding of the Src SH2 domain. This may stimulate Src activity by disrupting the interaction of the SH2 domain with pTyr 527 in the Src tail. *b*, The Src SH3 domain binds substrates. The SH3 domain of activated Src recognizes AFAP-110, which is then phosphorylated on tyrosine and associates with the SH2 domain.

turally similar to a *Drosophila* protein that functions as a negative regulator of Ras in the sevenless signalling pathway[93]. p120 GAP possesses SH2, SH3 and PH domains, and binds to pTyr-containing proteins, including the β-PDGFR. A phosphoprotein (p190) that recognizes the N-terminal SH2/SH3 region of p120 GAP in growth-factor-stimulated cells is itself a GAP for the Rho GTPases involved in organizing the actin cytoskeleton[95-97]. Control of gene expression by the Ras pathway and control of cell shape and adhesion by the Rho pathway may therefore be linked and coordinated through the association of the respective GAPs.

## Src SH2 and SH3 domains

The Ras pathway demonstrates that SH2 and SH3 domains not only work in concert, but can also have multiple binding partners (and hence several discrete functions). Similarly, the SH2 and SH3 domains of the cytoplasmic PTK Src regulate kinase activity, bind substrates both before and after phosphorylation by the kinase domain, and direct subcellular localization.

**SH2 and SH3 as regulators of Src.** Src (like Lck and other related proteins) has an SH2 and an SH3 domain adjacent to its kinase domain. Kinase activity is repressed when Tyr 527, within the short C-terminal tail, is phosphorylated by Csk, another intracellular kinase, as long as both the SH2 and SH3 domains are intact[98]. Phosphorylation at Tyr 527 thus appears to create a binding site for the Src SH2 domain, an event that also allows the SH3 domain to contact another part of Src. Together, these interactions prevent the SH2 and SH3 domains from binding to heterologous cellular proteins and lock the kinase domain in an inactive conformation. These interactions are generally thought to be intramolecular, largely because Src appears to be monomeric. Biophysical evidence indicates that an intramolecular SH2–pTyr interaction occurs in the Crk adaptor protein[99], suggesting that the same is true of Src. But the structure of the Lck SH2 and SH3 domains complexed with a phosphopeptide resembling the C-terminal Lck tail indicates that the inactive form of Lck (and by extension, Src) may be a dimer in which the SH2 and SH3 domains of opposing subunits are associated[100]. Irrespective of whether their interactions are inter- or intramolecular, however, the Src SH2 and SH3 domains clearly have an autoinhibitory role.

Two mechanisms can thus be envisaged for Src activation: either dephosphorylation of Tyr 527 destabilizes the complex by releasing the SH2 and SH3 domains and activating the kinase domain, or high-affinity ligands for the SH2 or SH3 domains compete for the autoinhibitory interactions, with the same effect. For example, autophosphorylation of the β-PDGFR in the juxtamembrane region creates a high-affinity binding site for the SH2 domains of Src and its relatives Fyn and Yes, whose kinase activities are stimulated on association with the receptor[101,102]. Similarly, autophosphorylation of the focal adhesion kinase FAK, a PTK activated by the adhesion of cell-surface integrins to extracellular fibronectin, creates a binding site for the Src SH2 domain[103], linking cell adhesion to the phosphorylation of Src substrates (Fig. 5*a*).

**Src SH2 and SH3 domains in substrate recognition.** Once Src is activated, its SH3 domain plays a direct role in the recognition of substrates. In cells transformed by oncogenic Src variants, an actin-associated protein of $M_r$ 110K containing proline-rich motifs (AFAP-110) is recognized by the Src SH3 domain, and phosphorylated by the kinase domain (Fig. 5*b*)[104]. Tyrosine phosphorylation of AFAP-110 cements its interaction with Src by creating a binding site for the Src SH2 domain, and may target Src to actin stress fibres.

Another substrate that may interact with Src family kinases in this way is the 68K protein Sam68 (ref. 105). Sam68 binds to Src and is phosphorylated on tyrosine during mitosis, when Src itself is activated[106,107]. Sam68 may be sequestered in the nucleus of interphase cells, and only be able to interact with c-Src on breakdown of the nuclear envelope. During mitosis, Sam68 binds through one of its proline-rich motifs to the Src SH3 domain, subsequently becomes phosphorylated and also engages the Src SH2 domain, and may thereby mediate a mitotic-specific signal from Src[108,109]. The nature of this signal is unknown, but as Sam68 contains a domain found in a variety of RNA-binding proteins[108,109] and binds RNA *in vitro*, it may regulate gene expression.

The Src SH2 domain can also bind substrates without the aid of the SH3 domain. The prominent pTyr-containing protein $p130^{cas}$, for example, associates specifically with the SH2 domain of activated Src *in vivo*[110]. This protein is also tyrosine-phosphorylated in v-*crk*-transformed cells, and becomes tightly bound to the v-Crk SH2 domain, indicating that it may stimulate mitogenesis[2]. It contains a remarkable series of motifs: an N-terminal SH3 domain is followed by 15 potential phosphorylation sites, with the consensus Tyr-X-X-Pro[111]. Nine of these closely spaced motifs have the sequence Tyr-Asp-Val/Thr-Pro, which conforms reasonably well to the consensus sequence recognized by the Crk SH2 domain. Like IRS-1, $p130^{cas}$ may therefore function as a docking protein, binding SH2-containing polypeptides through its pTyr sites. Its association with the Src SH2 domain may link it to the membrane, or may facilitate repeated phosphorylation by holding the substrate close to the kinase active site. The substrate specificities of cytoplasmic tyrosine kinases like Src, Fps and Abl generally match the binding preferences of their SH2 domains closely[131], indicating that the two may well have co-evolved to allow the kinase domain to create sites recognized by its SH2 partner.

The Src SH2 and SH3 domains thus have several distinct ligands each, and can form a series of complexes with distinct functions. The choice of binding partner is regulated by Src phosphorylation, and by the location and phosphorylation state of potential binding proteins. The ability of SH2 and SH3 domains to form sequential complexes with different partners may be general, greatly expanding the scope and versatility of intracellular signalling networks. A challenge for the future will be to establish the biological significance of these complexes.

## Location of signalling proteins

SH2, SH3 and PH domains all appear to control the location of signalling proteins within the cell. This is a particularly striking feature of SH3 domains, which are closely tied to the control of cell morphology. Several proteins associated with the cytoskeleton, including α-spectrin and myosin-1, possess SH3 domains. Genetic analysis in yeast has also demonstrated that SH3-containing proteins such as ABP-1, SLA1 and BEM1 are required for organization or polarization of the actin cytoskeleton[19,112,113]. SH3-containing proteins are also important for cellular organization in *Drosophila*, where mutations in the tumour-suppressor gene *discs large* (*dlg*) lead to a loss of the tight (septate) junctions between epithelial cells and aberrant proliferation of cells in the imaginal disk[114]. The Dlg protein has a C-terminal guanylate-kinase-like domain, a central SH3 domain, and a repeated N-terminal motif. The SH3 domain is important in maintaining the cellular junction, as a substitution in this domain results in loss of function (D. Woods and P. Bryant, personal communication), and may therefore anchor junctional components to the cytoskeleton.

SH3 domains are also important for subcellular localization and cytoskeletal interactions in mammalian cells. In mammalian fibroblasts, the SH3 domains of Grb2 and PLC-γ1 direct these proteins to membrane ruffles and actin stress fibres, respectively, indicating that their interactions control protein distribution within the cell[115]. Similarly, the SH3 domain of the cytoskeletal protein α-spectrin binds both *in vitro* and *in vivo* to a proline-rich motif in the cytoplasmic tail of an amiloride-sensitive rat sodium channel[116]. This interaction is apparently sufficient to direct the channel to the apical surface of epithelial cells.

An especially intriguing site of SH3 action is the NADPH oxidase system of phagocytic cells such as neutrophils[117], which

is activated by inflammatory stimuli to produce superoxide, a precursor to antimicrobial oxidants. The activated oxidase complex contains a cytochrome $b_{558}$ composed of two transmembrane proteins (gp91-phox and p22-phox), three proteins bearing SH3 domains (p47-phox and p67-phox, each with two SH3 domains, and p40-phox, with one), and a small GTP-binding protein, Rac. p47- and p67-phox are cytosolic in unstimulated cells, but move to the membrane and associate with the cytochrome $b_{558}$ in cells stimulated by phagocytosis or inflammatory mediators. Mutations in the genes for the gp91-, p22-, p47- or p67-phox components can cause chronic granulomatous disease, a condition characterized by high susceptibility to bacterial and fungal infections.

The SH3 domains of p47-phox and p67-phox may be responsible for the interactions that assemble the functional oxidase[118-120]. Recent work indicates that before stimulation, p47-phox may be in a closed conformation, with its SH3 domains engaging internal proline motifs. On activation, the C-terminal SH3 domain of p67-phos could bind specifically to the proline motif in the tail of p47-phox, while the SH3 domains of p47-phos could bind a proline-rich element in the membrane-bound p22-phox. These interactions are required for enzyme activation in intact cells. The Pro-to-Gln mutation in the SH3-binding motif of p22-phos detected in one patient with chronic granulomatons disease blocks the interaction of p22-phox with the p47-phox SH3 domains. Stimulation of phagocytic cells thus appears to activate a signalling pathway (as yet poorly defined) directed at the p47-phox and p67-phox SH3 domains, which induces a conformational change, reorganizing their SH3 ligands and recruiting the two proteins into the oxidase complex. Their function once bound to the cytochrome is unclear, although they may enhance the ability of gp91-phox to transport electrons from NADPH to $O_2$. The molecular events initiating the rearrangement of the SH3 ligands in p47-phox and p67-phox are also mysterious, although the C-terminal phosphorylation of p47-phox induced by neutrophil activation may help to break the interaction of its SH3 domain with the C-terminal proline-rich sequence. In the same way that the SH2 domains of Zap-70 mediate interactions important for T-cell activation, the SH3 domains of p47-phox and p67-phox are required for the response of phagocytic cells to pathogens.

## Summary

Numerous proteins involved in signal transduction possess SH2, SH3 and PH domains, either alone or in combination. Similarly, there are many polypeptides with binding sites, such as proline motifs or tyrosine phosphorylation sites, for these modules. The resulting interactions are important for signalling from the cell surface to the nucleus, for protein trafficking and subcellular localization, for the control of cell architecture and cell–cell interactions, and for cellular responses to infection. Genetic analysis has hinted at the biological functions of these domains. Biochemical experiments now indicate that they mediate an extremely complex and versatile network of interactions which may co-ordinate cellular responses to the environment.

The use of binding modules appears to be very general in signal transduction and in control of the cell cycle. In addition to the SH2, SH3 and PH domains, recently discovered modules include the PTB domain; the LIM domain, a cysteine-rich, zinc-binding unit which appears to regulate protein–protein interaction in a variety of transcription factors, cytoskeletal proteins and other signalling polypeptides[121]; the armadillo (arm) domain, a common element of proteins such as $\beta$-catenin which form part of cellular junctions and are involved in both cell adhesion and in cell–cell signalling[122]; and the Notch/ankyrin repeat, which is present in transmembrane receptors, cytoskeletal proteins and transcription factors, and mediates protein:protein interactions crucial for intercellular signalling[123]. Remarkably, a cytoplasmic tyrosine kinase similar to Zap-70 has been described in the simple metazoan *Hydra vulgaris*, in which the two SH2 domains flank five ankyrin repeats[124]. The plot thickens. □

*Tony Pawson is in the Division of Molecular and Developmental Biology, Samuel Lunenfeld Research Institute, Mount Sinai Hospital, 600 University Avenue, Toronto, Ontario M5G 1X5, Canada.*


1. Sadowski, I., Stone, J. C. & Pawson, T. *Molec. cell. Biol.* **6**, 4396–4408 (1986).
2. Mayer, B. J., Hamaguchi, M. & Hanafusa, H. *Nature* **332**, 272–275 (1988).
3. Tyers, M. *et al. Nature* **333**, 470–473 (1988).
4. Pawson, T. & Schlessinger, J. *Curr. Biol.* **3**, 434–442 (1993).
5. Anderson, D. *et al. Science* **250**, 979–982 (1990).
6. Matsuda, M., Mayer, B. J., Fukui, Y. & Hanafusa, H. *Science* **248**, 1537–1539 (1990).
7. Valius, M. & Kazlauskas, A. *Cell* **73**, 321–334 (1993).
8. Piccione, E. *et al. Biochemistry* **32**, 3197–3202 (1993).
9. Felder, S. *et al. Molec. cell. Biol.* **13**, 1449–1455 (1993).
10. Panayotou, G. *et al. Molec. cell. Biol.* **13**, 3567–3576 (1993).
11. Waksman, G., Shoelson, S., Pant, N., Cowburn, D. & Kuriyan, J. *Cell* **72**, 779–790 (1993).
12. Eck, M. I., Shoelson, S. E. & Harrison, S. C. *Nature* **362**, 87–91 (1993).
13. Pascal, S. M. *et al. Cell* **77**, 461–472 (1994).
14. Lee, C-H. *et al. Structure* **2**, 423–438 (1994).
15. Songyang, Z. *et al. Cell* **72**, 767–778 (1993).
16. Sugimoto, S., Wandlers, T. J., Shoelson, S. E., Neel, B. G. & Walsh, C. T. *J. biol. Chem.* **269**, 13614–13622 (1994).
17. Sabe, H., Hata, A., Okada, M., Nakagawa, H. & Hanafusa, H. *Proc. natn. Acad. Sci. U.S.A.* **91**, 984–3988 (1994).
18. Rotin, D., Honegger, A. M., Margolis, B. L., Ullrich, A. & Schlessinger, J. *J. biol. Chem.* **267**, 9678–9683 (1992).
19. Kavanaugh, W. M. & Williams, L. T. *Science*, **266**, 1862–1865 (1994).
20. Blaike, P. *et al. J. biol. Chem.* **269**, 32031–32034 (1994).
21. Drubin, D. G., Mulholland, J., Zhu, Z. & Botstein, D. *Nature* **343**, 288–290 (1990).
22. Leto, T. L. *et al. Science* **248**, 727–730 (1990).
23. Ren, R., Mayer, B. J., Cicchetti, P. & Baltimore, D. *Science* **259**, 1157–1161 (1993).
24. Yu, H. *et al. Cell* **76**, 933–945 (1994).
25. Chen, J. K., Lane, W. S., Brauer, A. W., Tanaka, A. & Schreiber, S. L. *J. Am. chem. Soc.* **115**, 12591–12592 (1993).
26. Feng, S., Chen, J. K., Yu, H., Simon, J. A. & Schreiber, S. L. *Science* **266**, 1241–1246 (1994).
27. Lim, W. A., Richards, F. M. & Fox, R. O. *Nature* **372**, 375–379 (1994).
28. Musacchio, A., Saraste, M. & Wilmanns, M. *Nature struct. Biol.* **1**, 546–551 (1994).
29. Wittekind, M. *et al. Biochemistry* **33**, 13531–13539 (1994).
30. Rickles, R. *et al. EMBO J.* **13**, 5598–5604 (1994).
31. Cherniack, A. D., Klarlund, J. K. & Czech, M. P. *J. biol. Chem.* **269**, 4717–4720 (1994).
32. Mayer, B. J., Ren, R., Clark, K. L. & Baltimore, D. *Cell* **73**, 629–630 (1993).
33. Haslam, R. J., Koide, H. B. & Hemmings, B. A. *Nature* **363**, 309–310 (1993).
34. Musacchio, A., Gibson, T., Price, P., Thompson, J. & Saraste, M. *Trends biochem. Sci.* **18**, 343–348 (1993).
35. Riddihough, G. *Nature struct. Biol.* **1**, 755–757 (1994).
36. Harlan, J. E., Hajduk, P. J., Yoon, H. S. & Fesick, S. W. *Nature* **371**, 168–170 (1994).
37. Ferguson, K. M., Lemmon, M. A., Schlessinger, J. & Sigler, P. B. *Cell* **79**, 199–209 (1994).
38. Timm, D. *et al. Nature struct. Biol.* **1**, 782–788 (1994).
39. Touhara, K., Inglese, J., Pitcher, J. A., Shaw, G. & Lefkowitz, R. J. *J. biol. Chem.* **269**, 10217–10220 (1994).
40. Yao, L., Kawakami, Y. & Kawakami, T. *Proc. natn. Acad. Sci. U.S.A.* **91**, 9175–9179 (1994).
41. Lefkowitz, R. L. *Cell*, **74**, 409–412 (1993).
42. Thomas, J. D. *et al. Science* **261**, 355–358 (1993).
43. Ullrich, A. & Schlessinger, J. *Cell* **61**, 203–212 (1990).
44. Rönnstrand, L. *et al. EMBO J.* **11**, 3911–3919 (1992).
45. Valius, M., Bazenet, C. & Kazlauskas, A. *Molec. cell. Biol.* **13**, 133–143 (1993).
46. Kashishian, A. & Cooper, J. A. *Molec. Biol. Cell.* **4**, 49–57 (1993).
47. Fantl, W. J. *et al. Cell* **69**, 413–423 (1992).
48. Twamley-Stein, G. M., Pepperkok, R., Ansorge, W. & Courtneidge, S. A. *Proc. natn. Acad. Sci. U.S.A.* **90**, 7696–7700 (1993).
49. Satoh, T., Fantl, W. J., Escobedo, J. A., Williams, L. T. & Kaziro, Y. *Molec. cell. Biol.* **13**, 3706–3713 (1993).
50. Ponzetto, C. *et al. Cell* **77**, 261–271 (1994).
51. Sun, X. J. *et al. Nature* **352**, 73–77 (1991).
52. Sun, X. J., Crimmins, D. L., Myers, M. G. Jr, Miralpeix, M. & White, M. F. *Molec. cell. Biol.* **13**, 7418–7428 (1993).
53. Reth, M. *Nature* **338**, 383–384 (1989).
54. Weiss, A. *Cell* **73**, 209–212 (1993).
55. Iwashima, M., Irving, B. A., van Oers, N. S. C., Chan, A. C. & Weiss, A. *Science* **263**, 1136 (1994).
56. Arpaia, E., Shahar, M., Dadi, H., Cohen, A. & Roifman, C. M. *Cell* **76**, 947–958 (1994).
57. Chan, A. C. *et al. Science* **264**, 1599 (1994).
58. Prasad, K. V. S. *et al. Proc. natn. Acad. Sci. U.S.A.* **91**, 2834–2838 (1994).
59. Pagès, F. *et al. Nature* **369**, 327–329 (1994).
60. Darnell, J. E. Jr, Kerr, I. M. & Stark, G. R. *Science* **264**, 1415–1421 (1994).
61. Stahl, N. & Yancopoulos, G. D. *Cell* **74**, 587–590 (1993).
62. Watling, D. *et al. Nature* **368**, 166–170 (1993).
63. Müller, M. *et al. Nature* **366**, 129–135 (1993).
64. Maruyamam, K., Miyata, K. & Yoshimura, A. *J. biol. Chem.* **269**, 5976–5980 (1994).
65. Yi, T., Mui, A. L.-F., Krystal, G. & Ihle, J. N. *Molec. cell. Biol.* **13**, 7577–7586 (1993).
66. Boulton, T. G., Stahl, N. & Yancopoulos, G. D. *J. biol. Chem.* **269**, 11648–11655 (1994).
67. Schindler, C., Fu, X-Y., Improta, T., Aeborsold, R. & Darnell, J. E. *Proc. natn. Acad. Sci. U.S.A.* **89**, 7836–7839 (1992).
68. Fu, X.-Y. *Cell* **70**, 323–335 (1992).
69. Shuai, K., Stark, G. R., Kerr, I. M. & Darnell, J. E. *Science* **261**, 1744–1746 (1993).
70. Shuai, K. *et al. Cell* **76**, 821–828 (1994).
71. Clark, S. G., Stern, M. J. & Horvitz, H. R. *Nature* **356**, 340–344 (1992).
72. Simon, L. A., Dodson, G. S. & Rubin, G. M. *Cell* **73**, 169–177 (1993).
73. Olivier, J. P. *et al. Cell* **73**, 179–191 (1993).
74. Rozakis-Adcock, M., Fernley, R., Wade, J., Pawson, T. & Bowtell, D. *Nature* **363**, 83–85 (1993).

75. Egan, S. E. *et al. Nature* **363,** 45–51 (1993).
76. Li, N. *et al. Nature* **363,** 85–88 (1993).
77. Gale, W. N., Kaplan, D., Lowenstein, E. J., Schlessinger, J. & Bar-Sagi, D. *Nature* **363,** 88–92 (1993).
78. Buday, L. & Downward, J. *Cell* **73,** 611–620 (1993).
79. Batzer, A. G., Rotin, D., Urena, J. M., Skolnik, E. Y. & Schlessinger, J. *Molec. cell. Biol.* **14,** 5192–5201 (1994).
80. Quilliam, L. A. *et al. Proc. natn. Acad. Sci. U.S.A.* **91,** 8512–8516 (1994).
81. Aronheim, A. *et al. Cell* **78,** 949–961 (1994).
82. Marshall, C. *Curr. Opin. Genet. Dev.* **4,** 82–89 (1994).
83. Arvidsson, A-K. *et al. Molec. cell. Biol.* **14,** 6715–6726 (1994).
84. Rozakis-Adcock, M. *et al. Nature* **360,** 689–692 (1992).
85. Li, W. *et al. Molec. cell. Biol.* **14,** 509–517 (1994).
86. Bennett, A. M., Tang, T. L., Sugimoto, S., Walsh, C. T. & Neel, B. G. *Proc. natn. Acad. Sci. U.S.A.* **91,** 7335–7339 (1994).
87. Miki, H. *et al. J. biol. Chem.* **269,** 5489–5492 (1994).
88. Scaife, R., Gout, I., Waterfield, M. D. & Margolis, R. L. *EMBO J.* **13,** 2574–2582 (1994).
89. Seedorf, K. *et al. J. biol. Chem.* **269,** 16009–16014 (1994).
90. Tanaka, S. *et al. Proc. natn. Acad. Sci. U.S.A.* **91,** 3443–3447 (1994).
91. Matsuda, M. *et al. Molec. cell. Biol.* **14,** 5495–5500 (1994).
92. Boguski, M. S. & McCormick, F. *Nature* **366,** 643–654 (1993).
93. Maekawa, M. *et al. Molec. cell. Biol.* **14,** 6879–6885 (1994).
94. Downward, J., Graves, J. D., Warne, P. H., Rayter, S. & Cantrell, D. A. *Nature* **346,** 719–723 (1990).
95. Settleman, J., Albright, C. F., Foster, L. C. & Weinberg, R. A. *Nature* **359,** 153–154 (1992).
96. McGlade, J. *et al. EMBO J.* **12,** 3073–3081 (1993).
97. Nobes, C. & Hall, A. *Curr. Opin. Genet. Dev.* **4,** 77–81 (1994).
98. Superti-Furga, G., Fumagalli, S., Koegl, M., Courtneidge, S. A. & Draetta, G. *EMBO J.* **12,** 2625–2634 (1993).
99. Rosen, M. K. *et al. Nature* (in the press).
100. Eck, M. J., Atwell, S. K., Shoelson, S. E. & Harrison, S. C. *Nature* **368,** 764–769 (1994).
101. Kypta, R. M., Goldberg, Y., Ulug, E. T. & Courtneidge, S. A. *Cell* **62,** 481–492 (1990).
102. Mori, S. *et al. EMBO J.* **12,** 2257–2264 (1993).
103. Cobb, B. S., Schaller, M. D., Leu, T.-H. & Parsons, J. T. *Molec. cell. Biol.* **14,** 147–155 (1994).
104. Flynn, D. C., Leu, T-H., Reynolds, A. B. & Parsons, T. *Molec. cell. Biol.* **13,** 7892–7900 (1993).
105. Courtneidge, S. A. & Fumagalli, S. *Trends Cell Biol.* **4,** 345–347 (1994).
106. Taylor, S. J. & Shalloway, D. *Nature* **368,** 867–871 (1994).
107. Fumagalli, S., Totty, N. F., Hsuan, J. J. & Courtneidge, S. A. *Nature* **368,** 871–874 (1994).
108. Wong, G. *et al. Cell* **69,** 551–558 (1992).
109. Richard, S. *et al. Molec. cell. Biol.* (in the press).
110. Reynolds, A. B., Kanner, S. B., Wang, H.-C. R. & Parson, J. T. *Molec. cell. Biol.* **9,** 3951–3958 (1989).
111. Sakai, R. *et al. EMBO J.* **13,** 3748–3756 (1994).
112. Holtzman, D. A., Yang, S. & Drubin, D. G. *J. Cell Biol.* **122,** 635–644 (1993).
113. Chenevert, J., Corrado, K., Bender, A., Pringle, J. R. & Herskowitz, I. *Nature* **356,** 77–79 (1992).
114. Woods, D. F. & Bryant, P. J. *Cell* **66,** 451–464 (1991).
115. Bar-Sagi, D., Rotin, D., Batzer, A., Mandiyan, V. & Schlessinger, J. *Cell* **74,** 83–91 (1993).
116. Rotin, D. *et al. EMBO J.* **13,** 4440–4450 (1994).
117. McPhail, L. C. *J. Exp. Med.* **180,** 2011–2015 (1994).
118. de Mendez, I., Garrett, M. C., Adams, A. G. & Leto, T. L. *J. biol. Chem.* **269,** 16326–16332 (1994).
119. Finan, P. *et al. J. biol. Chem.* **269,** 13752–13755 (1994).
120. Sumimoto, H. *et al. Proc. natn. Acad. Sci. U.S.A.* **91,** 5345–5349 (1994).
121. Schmeicel, K. L. & Beckerle, M. C. *Cell* **79,** 211–219 (1994).
122. Peifer, M., Berg, S. & Reynolds, A. B. *Cell* **76,** 789–791 (1994).
123. Fortini, M. F. & Artavanis-Tsakonas, S. *Cell* **75,** 1245–1247 (1993).
124. Chan, T. A. *et al. Oncogene* **9,** 1253–1259 (1993).
125. Perkins, L. A., Larsen, I. & Perrimon, N. *Cell* **70,** 225–236 (1992).
126. Feller, S. M., Knudsen, B. & Hanafusa, H. *EMBO J.* **13,** 2341–2351 (1994).
127. Ren, R., Ye, Z-S. & Baltimore, D. *Genes Dev.* **8,** 783–795 (1994).
128. Rodriguez-Viciana, P. *et al. Nature* **370,** 527–532 (1994).
129. Gustafson, T. A. *et al. Molec. cell. Biol.* (in the press).
130. Stahl, N. *et al. Science* (in the press).
131. Songyang, Z. *et al. Nature* **373,** 536–539 (1995).

ACKNOWLEDGEMENTS. I thank M. Rosen, P. van der Geer and M. Henkemeyer for critical comments on the manuscript. Apologies are due to the authors of the many important papers not cited or discussed here, or referenced indirectly, for want of space. The author is a Terry Fox Cancer Research Scientist of the National Cancer Institute of Canada.



# Crystal structure of the nickel–iron hydrogenase from *Desulfovibrio gigas*

**Anne Volbeda, Marie-Hélène Charon, Claudine Piras, E. Claude Hatchikian*, Michel Frey† & Juan C. Fontecilla-Camps†**

Laboratoire de Cristallographie et de Cristallogénèse des Protéines, Institut de Biologie Structurale J. P. Ebel (CEA,CNRS), 41 avenue des Martyrs, 38027 Grenoble Cedex 1, France
* Unité de Bioénergétique et Ingéniérie des Protéines, CNRS, 13402 Marseille Cedex 20, France

**The X-ray structure of the heterodimeric Ni–Fe hydrogenase from *Desulfovibrio gigas*, the enzyme responsible for the metabolism of molecular hydrogen, has been solved at 2.85 Å resolution. The active site, which appears to contain, besides nickel, a second metal ion, is buried in the 60K subunit. The 28K subunit, which coordinates one [3Fe–4S] and two [4Fe–4S] clusters, contains an amino-terminal domain with similarities to the redox protein flavodoxin. The structure suggests plausible electron and proton transfer pathways.**

THE capacity to oxidize molecular hydrogen or reduce protons ($H_2 \Leftrightarrow 2H^+ + 2e^-$) is a central metabolic feature of a wide variety of microorganisms[1]. $H_2$ oxidation is coupled to the reduction of electron acceptors such as oxygen, nitrate, sulphate, carbon dioxide and fumarate, whereas proton reduction ($H_2$ evolution) is an essential element of pyruvate fermentation or the disposal of excess electrons. A $H_2$-cycling process that would contribute to the generation of a proton gradient during growth on organic substrates in the presence of sulphate has been reported in some sulphate-reducing bacteria[2]. Three types of metal-containing hydrogenases, the enzymes responsible for hydrogen metabolism in microorganisms, have been defined: 'iron only' or Fe hydrogenases[3], Ni–Fe hydrogenases and Ni–Fe–Se hydrogenases[4–6].

† To whom correspondence should be addressed.

The Ni–Fe hydrogenase from *Desulfovibrio gigas* is a heterodimeric periplasmic protein consisting of 28K and 60K subunits[7,8]. The latter subunit matures to a functional form through cleavage of its 15 C-terminal residues[9]. Besides the nickel ion[10], *D. gigas* hydrogenase contains three iron–sulphur centres[7]. These centres have been identified as one [3Fe–4S] and two [4Fe–4S] clusters[11]. The enzyme isolated in aerobic conditions exhibits two Ni-related electron paramagnetic resonance (EPR) signals, called Ni-A and Ni-B[12,13]. The Ni-A signal corresponds to the so-called 'unready' state of hydrogenase whose complete activation may take several hours. The Ni-B signal corresponds to the 'ready' state, which can be quickly reduced to the active form by $H_2$ under anaerobic conditions thereby developing $H_2$-uptake activity in the presence of electron acceptors[12,13]. During reduction in the presence of hydrogen, the enzyme displays a light-sensitive EPR signal called Ni-C, which

*Proc. Natl. Acad. Sci. USA*
Vol. 87, pp. 31–35, January 1990
Biochemistry

# *Drosophila* glutathione *S*-transferase 1-1 shares a region of sequence homology with the maize glutathione *S*-transferase III

YANN-PYNG S. TOUNG*, TAO-SHIH HSIEH†, AND CHEN-PEI D. TU*‡

*Department of Molecular and Cell Biology, Pennsylvania State University, University Park, PA 16802; and †Department of Biochemistry, Duke University School of Medicine, Durham, NC 27709

*Communicated by James C. Wang, September 22, 1989 (received for review June 25, 1989)*

**ABSTRACT** **We have characterized a *Drosophila* glutathione *S*-transferase (RX:glutathione R-transferase, EC 2.5.1.18) cDNA encoding a protein of 209 amino acids. The cDNA was expressed in *Escherichia coli* harboring the expression plasmid construct pGTDml-KK. The active enzyme, designated as *Drosophila* glutathione *S*-transferase 1-1, had a specific activity toward 1-chloro-2,4-dinitrobenzene comparable to that for the mammalian glutathione *S*-transferases but did not have as broad a substrate specificity pattern. There is a region of 44 amino acids in this enzyme that shares 66% identity with an analogous region of maize glutathione *S*-transferase III. *Drosophila* glutathione *S*-transferase 1-1 had no obvious homology to any mammalian or parasitic glutathione *S*-transferases. The gene was found to be a member of a multigene family.**

The glutathione *S*-transferases (RX:glutathione R-transferases EC 2.5.1.18; GSTs) are a family of multifunctional proteins (for recent reviews, see refs. 1 and 2). High multiplicity of GSTs with overlapping substrate specificities may be essential to their multiple roles in xenobiotic metabolism, drug biotransformation, and protection against peroxidative damage. This isozyme is ubiquitous among eukaryotes. The rat and human GSTs are products of their respective gene superfamilies (3–6). In plants, GSTs are involved in the detoxification of certain herbicides. Different specificities of plant GSTs are thought to be the basis of selective actions of some herbicides (7–10). It has also been proposed that resistance to certain selected pesticides in insects may be related to changes in their GST expression (11, 12). The parasitic helminths of the genus *Schistosoma* have surface antigens that are glutathione *S*-transferases. Acquired immunity in mice, rats, hamsters, and monkeys against this antigen from *Schistosoma japonicum* or *Schistosoma mansoni* has provided protection against schistosomiasis, a chronic debilitating disease in several parts of the world (13–15). The major squid lens crystallins may be themselves GSTs or evolutionarily related to GSTs (16, 17). In this communication, we report the molecular characterization of a *Drosophila* GST gene[§] and its heterospecific expression in *Escherichia coli*.

## MATERIALS AND METHODS

**Materials.** Chemicals, *S*-hexylglutathione (GSH)-linked agarose and antibiotics were purchased from Sigma and/or Merck. Radioactive nucleotides were products of Amersham or DuPont/NEN. $^{125}$I-labeled protein A was purchased from ICN. Restriction endonucleases and T4 DNA ligase were products from New England Biolabs. A nick-translation kit was purchased from BRL. $Kc_0$ cells were grown in spinner flasks to a density of $\approx 5 \times 10^6$ cells per ml, according to a published procedure (18). Twelve- to 16-hour-old *Drosophila* embryos (Oregon R) were collected and washed before use. A *Drosophila* head cDNA library in the λgtll vector was provided by P. Salvaterra of the Beckman Research Institute (Duarte, CA) (19). A genomic library in the λEMBL4 vector was constructed from Oregon R genomic DNA as described (20).

**Purification of GSTs from $Kc_0$ Cells and *Drosophila* Embryos.** One unit of GST activity catalyzes the formation of 1-chloro-2,4-dinitrobenzene (CDNB)–GSH conjugate at a rate of 1 μmol/min. $Kc_0$ cells ($\approx 1 \times 10^{11}$) were lysed with Nonidet P-40 and cycles of freezing-and-thawing. After removal of nuclei by centrifugation at $5000 \times g$ (6000 rpm, SS34 rotor, Sorvall centrifuge), the supernatant fractions were brought to 70% saturation by adding solid ammonium sulfate. The precipitated proteins, which contained nearly all the GST activities determined by CDNB conjugation, were recovered by centrifugation ($8000 \times g$ for 20 min). The precipitate was dissolved in a minimum volume of 25 mM Tris·HCl (pH 8.0) and dialyzed against the same buffer (1 liter) overnight with one change. Traces of denatured proteins were removed from the dialyzed fraction by centrifugation before affinity chromatography on *S*-hexyl-GSH (3-ml bed volume, Econo-column from Bio-Rad). Approximately 35% of the CDNB-conjugation activities appeared in the flow-through fractions that did not bind to a second column of *S*-hexyl-GSH-linked agarose; these fractions were not processed further for the results reported here. The affinity column was washed with 25 mM Tris·HCl (pH 8.0) buffer that contained 0.2 M KCl, and GST activities were eluted according to published procedures (21, 22). The eluted fractions contained two major bands on SDS/PAGE: one band at 23.4 kDa and the other band at 28.5 kDa (data not shown). This enzyme sample was dialyzed against 50 mM 2-[*N*-morpholino]ethanesulfonic acid (MeS) (pH 6.1) and further purified by fast protein liquid chromatography on a Mono Q column. The elution was carried out with a gradient of 0–0.3 M KCl in MeS buffer (pH 6.1). The first activity peak (peak I), which appeared just before the beginning of the gradient, contained a single band (23.4 kDa) on SDS/PAGE. The rest of the GST activities were eluted between 60 mM and 140 mM KCl in two overlapping activity peaks, peak II and peak III. Fractions of peak III contained four bands between 23.4 kDa and 31 kDa on SDS/PAGE. Peak I and peak III GSTs were used separately to raise polyclonal antibodies in rabbits as described

The publication costs of this article were defrayed in part by page charge payment. This article must therefore be hereby marked "*advertisement*" in accordance with 18 U.S.C. §1734 solely to indicate this fact.

Abbreviations: GST, glutathione S-transferase; GSH, glutathione; CDNB, 1-chloro-2,4-dinitrobenzene.

‡To whom reprint requests should be addressed at: Department of Molecular and Cell Biology, 6 Althouse Laboratory, Pennsylvania State University, University Park, PA 16802.

§The sequence reported in this paper has been deposited in the GenBank data base (accession no. X14233).

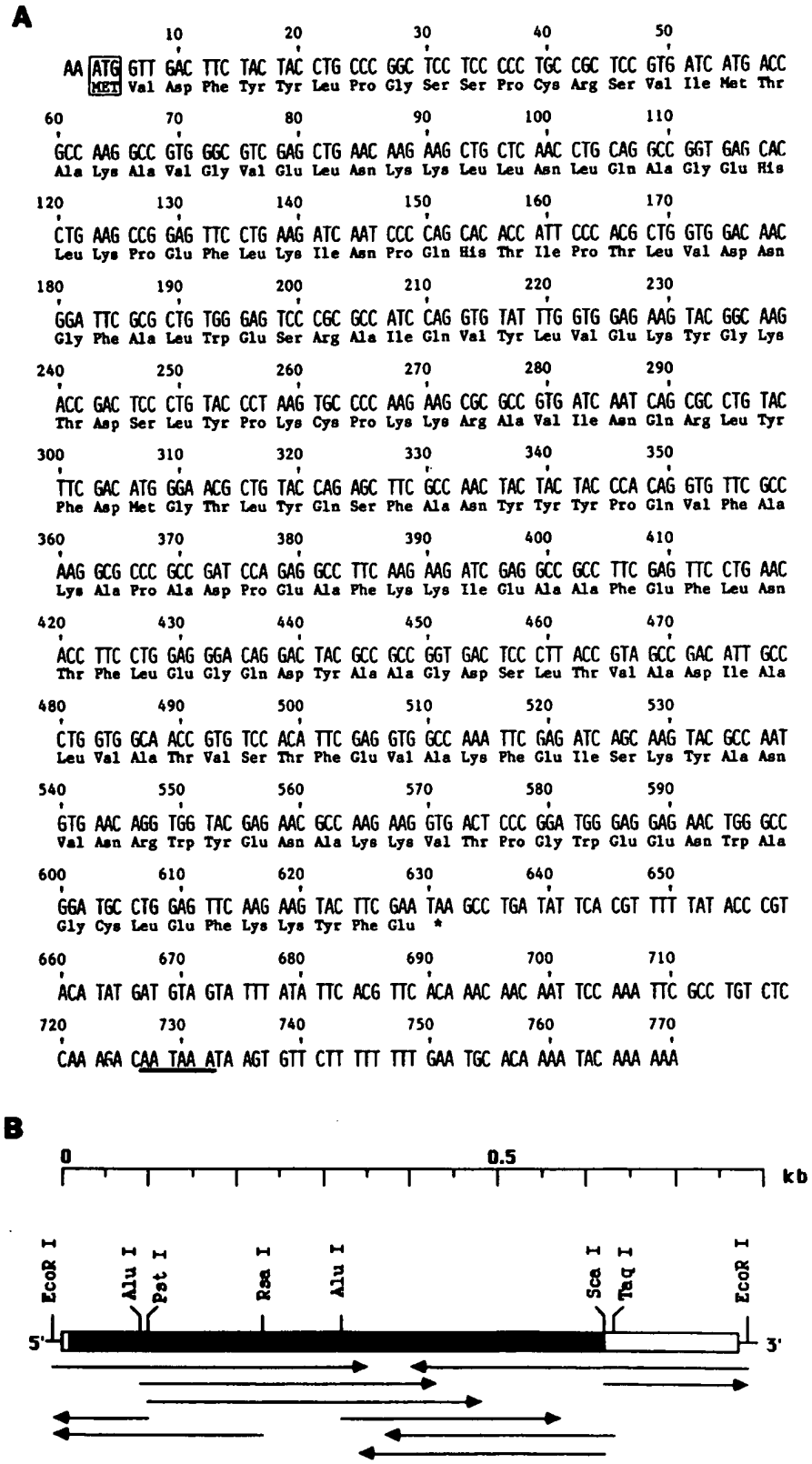



FIG. 1. Characterization of cDNA clone pGTDm1. (*A*) Nucleotide and deduced amino acid sequence. The ATG initiation codon is boxed, and the termination codon is indicated by an asterisk. The putative poly(A) addition signal is underlined. Arg-14, Pro-54, Asp-58, Ala-68, Ile-69, Gly-144, Asp-151, and Leu-160 are the eight residues identical to those conserved in mammalian GSTs (*5*, *6*). (*B*) Partial restriction map and sequencing strategy. Only restriction sites pertinent to subcloning into M13mp18/19 phage are shown. Arrows and lines represent the direction and extent of the sequence determinations. The protein coding region is shown by the solid black bar.

(21, 23). Specificity of the two antisera was tested by Western (immunological) blotting experiments (24) using fast protein liquid chromatography-purified *Drosophila* GSTs.

GSTs were also purified from *Drosophila* embryos. Embryo cytoplasm (≈460 ml) was concentrated by $(NH_4)_2SO_4$ precipitation (70% saturation) and dialyzed against 25 mM Tris·HCl (pH 8.0) before affinity chromatography. The remaining procedures were similar to those in the purification of GSTs from $Kc_0$ cells. Peak I fractions, which were eluted from the Mono Q column just before the beginning of the gradient, were subjected to N-terminal sequencing for 20 cycles with an Applied Biosystems model 477A microsequenator.

**Screening of *Drosophila* λgt11 cDNA Library with Anti-GST Antisera.** The *Drosophila* cDNA library was screened as described (25, 26) with $^{125}I$-labeled protein A to label antibody-positive plaques on nitrocellulose filters. The antibody against the peak III GST did not produce any useful clones. Two positive signals were detected among 100,000 plaques with antibody against peak I GST. The two positive clones were designated λGTDm1 and λGTDm1-1; their subclones in pUC19 were designated as pGTDm1 and pGTDm1-1, respectively.

**Screening of λEMBL4 Genomic DNA Library.** The cDNA insert in pGTDm1 used as a hybridization probe was purified twice from a polyacrylamide gel (20). Three positive clones were then isolated from screening a total of 50,000 plaques. DNAs from two of these clones were identical in restriction patterns, and they were designated λGTDm101. The third clone was designated λGTDm102.

**Methods for Nucleic Acid Analysis.** Plasmid and phage DNA isolation procedures were those described (20), except that phage DNAs were further purified by KI gradient centrifugation (27). *Drosophila* genomic DNA was purified from Oregon R embryos by a published procedure (20). DNA fragments were subcloned into plasmid (pKK223-3 or pUC19) and M13 phage (mp 18, mp 19) vectors as described (20). Southern hybridization (genomic or cloned DNA) was accomplished on nitrocellulose filters with the full-length cDNA insert or a DNA fragment containing just the 3′-noncoding region (28). DNA sequence was analyzed from M13 subclones by the dideoxynucleotide chain-termination method with [$\alpha$-$^{35}S$]dATP (29–31).

**Heterospecific Expression of a *Drosophila* GST cDNA in *E. coli*.** The cDNA insert of pGTDm1 was cloned into the *Eco*RI site of pKK223-3 in two orientations in *E. coli* JM105. The plasmid construct of correct orientation was designated as pGTDm1-KK, which expressed high levels of GST activity upon induction of the *tac* promoter by isopropyl β-D-thiogalactoside. The insert in the opposite orientation was designated as pGTDm1-KK′. To purify GST expressed from pGTDm1-KK, the procedures previously reported for the purification of *E. coli*-expressed $H_a$ subunit cDNAs and $Y_{b2}$ cDNA were followed up to the affinity chromatography step (32, 33). The proteins eluted from the *S*-hexyl-GSH affinity column were dialyzed thoroughly against 25 mM Tris·HCl (pH 8.0) and then separated by ion-exchange chromatography on a fast protein liquid chromatography Mono Q column. Enzyme activity was eluted in a single peak with a linear NaCl gradient (0–1 M) in 25 mM Tris·HCl (pH 8.0). The purified GST was dialyzed against 500 vol of 10 mM Tris·HCl (pH 8.0) and stored in aliquots at −20°C. Protein concentration was determined by the bincinchonic acid (BCA) protein assay (Pierce) with bovine serum albumin as standard.

## RESULTS

**Characterization of pGTDm1 and pGTDm1-1 Clones.** DNAs were isolated from λGTDm1 and λGTDm1-1 and analyzed after *Eco*RI digestion. Each clone contained an insert DNA of ≈750 base pairs (bp) in size. These clones were subcloned into M13mp18 and -mp19 vectors for sequence analysis either directly or after further digestion according to the strategy in Fig. 1. The two cDNAs were identical in sequence except that the pGTDm1-1 insert (data not shown) did not have six adenylate residues at the 3′ end (Fig. 1). The open reading frame of 209 amino acids had a calculated $M_r$ of 23,839. The deduced N-terminal sequence Met-Val-Asp-Phe-Tyr-Tyr-Leu-Pro-Gly-Ser-Ser-Pro-Cys-Arg-Ser-Val. . . was consistent with the protein sequence determined from peak I GST purified from $Kc_0$ cells or *Drosophila* embryos (Xaa-Xaa-Asp-Phe-(Ser)-Tyr-Leu-Pro-Gly-Ser-Xaa-Pro-(Trp)-Xaa-Ser-Xaa-Ile-Met-Thr-Ala-Xaa). The sequence match at the N-terminal region, the size of the open reading frame, and its positive identification by antibody against purified *Drosophila* GST make pGTDm1 a probable candidate for a *Drosophila* GST cDNA clone.

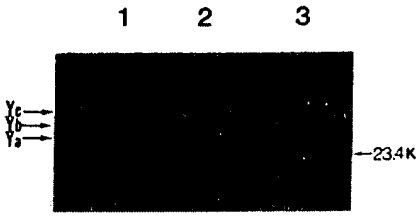



FIG. 2. Electrophoretic analysis of *Drosophila* GST 1-1. Lanes: 1, rat liver GSTs (≈2 μg) (Sigma); 2, *E. coli*-expressed *Drosophila* GST 1-1 (≈1 μg); 3, GST purified from *Drosophila* embryos (≈1 μg). Both *E. coli*-expressed and embryonic GSTs were purified by *S*-hexyl-GSH affinity chromatography and Mono Q column chromatography on fast protein liquid chromatography. The gel was silver-stained by using a Bio-Rad silver stain kit. The GST subunit 1 from *Drosophila* had an electrophoretic mobility at $M_r$ = 23,400. Rat liver GST $Y_a$, $Y_b$, and $Y_c$ subunits have $M_r$ values = 25,600, 27,000, and 28,000, respectively (1, 2).

**Heterospecific Expression of pGTDm1 cDNA in *E. coli*.** The two constructs in *E. coli* JM105 in opposing orientations, pGTDm1-KK and pGTDm1-KK′, were induced separately by isopropyl β-D-thiogalactoside (32, 33). The GST activity against CDNB could be detected only in sonicated extracts of cultures of pGTDm1-KK at a level of 0.09 unit per mg of protein. After purification of this activity from large-scale induced cultures (4-liter cultures) by affinity chromatography and fast protein liquid chromatography Mono Q column, we recovered the activity (17% yield) in a fraction containing a single band on SDS/PAGE. Its mobility (23.4 kDa) was identical to that of the purified embryo GST (Fig. 2). Thus, we have proven that the pGTDm1 cDNA insert encodes a *Drosophila* GST subunit, designated as subunit 1. When the fast protein liquid chromatography-purified GST (both *E. coli*-expressed and $Kc_0$ cell peak I enzyme) was chromatographed on a Sephadex G-100 column, CDNB-conjugation activities appeared at elution volumes of $M_r$ 48,500 relative to marker proteins (aprotinin, cytochrome *c*, carbonic anhydrase, bovine serum albumin). Therefore, GST subunit 1

Table 1. Substrate specificities of *E. coli*-expressed *Drosophila* GST 1-1

| Substrate | Specific activity, μmol/min per mg |
|---|---|
| 1-Chloro-2,4-dinitrobenzene | 24.08 ± 2.23* |
| Ethacrynic acid | 0.24 |
| 1,2-Dichloro-4-nitrobenzene | <0.01 |
| *p*-Nitrobenzyl chloride | <0.01 |
| 1,2-Epoxy-3-(*p*-nitrophenoxy)propane | <0.01 |
| Cumene hydroperoxide | <0.01 |

*Average of three determinations.

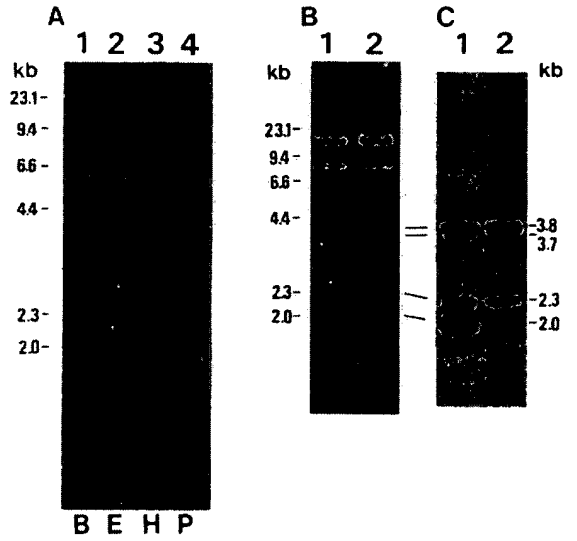



FIG. 3. Southern analysis of *Drosophila* genomic DNA and two *Drosophila* genomic clones, λGTDm101 and λGTDm102. (*A*) *Drosophila* genomic DNA (≈2.6 μg each) was digested overnight with 200 units each of *Bam*HI (B) (lane 1), *Eco*RI (E) (lane 2), *Hin*dIII (H) (lane 3), and *Pst* I (P) (lane 4) and separated by gel (0.8% agarose) electrophoresis. After transfer to a nitrocellulose membrane, the DNA was hybridized with the $^{32}$P-labeled *Eco*RI insert of pGTDm1. (*B*) *Eco*RI restriction pattern of two genomic clones, λGTDm101 (≈1 μg) (lane 1) and λGTDm102 (≈1 μg) (lane 2). (*C*) Autoradiogram of Southern analysis of λGTDm101 (lane 1) and λGTDm102 (lane 2) revealed by the pGTDm1 cDNA probe. Numbers at left of *A* and *B* indicate size markers (λ-*Hin*dIII) in kb. Numbers at right of *C* indicate sizes of positive DNA fragments. Lines between *B* and *C* indicate corresponding *Eco*RI fragments. Hybridizations were done in 5× SSC (1× SSC is 0.15 M sodium chloride/0.015 M sodium citrate, pH 7)/5× Denhardt's solution (1× Denhardt's solution is 0.02% polyvinylpyrrolidone/0.02% Ficoll/0.02% bovine serum albumin)/50 mM sodium pyrophosphate/0.1% NaDodSO$_4$/denatured salmon testis DNA at 100 μg/ml, 40% formamide (vol/vol) at 40°C for 40 hr. The cDNA probe had a specific activity of 2 × 10$^8$ dpm/μg of DNA. The filters were washed at room temperature three times in 2× SSC/0.1% NaDodSO$_4$ for 90 min and then washed in 0.1× SSC/0.1% NaDodSO$_4$ at 55°C for 2 hr.

most probably exists as active dimers, designated as *Drosophila* GST 1-1. The substrate specificity of GST 1-1 is described in Table 1. CDNB was the most active substrate, but this protein (at 19.6–392 nM enzyme concentration) did not have any detectable GSH peroxidase activity against cumene hydroperoxide relative to a similar concentration of human GST 1-1 (33).

**Genomic Complexity Detected by pGTDm1 cDNA.** Genomic Southern blot analysis with pGTDm1 cDNA insert as a probe revealed multiple bands upon various restriction endonuclease digestions (Fig. 3). The 2.0-kilobase (kb) fragment in λGTDm101 was also strongly positive toward the 3′-noncoding probe of pGTDm1 (data not shown). In the *Eco*RI-digested sample, the 2.8-kb band has not been cloned in the two genomic clones λGTDm101 and λGTDm102. Other restriction digestions revealed multiple bands >4 kb in size (Fig. 3*A*).

## DISCUSSION

We have isolated and characterized a cDNA encoding the *Drosophila* GST subunit 1 of $M_r$ = 23,839, which contains 209 amino acids. The CDNB conjugation activity (24 μmol/min per mg of protein) of *Drosophila* GST 1-1 is comparable to that for mammalian GSTs and is approximately twice the reactivity of heterodimeric GST of *Drosophila* larvae and adults, as characterized by Cochrane *et al.* (34). The *Drosophila* GST did not have as broad a substrate specificity pattern as most mammalian GSTs relative to the substrates tested in Table 1. The antibody against peak I GST does not cross-react with other *Drosophila* GSTs. It is clear that *Drosophila* GST is an isozyme family composed of at least three classes of subunits of 23.4 kDa, 28.5 kDa, and 35 kDa (ref. 34 and this study).

The size of *Drosophila* GST subunit 1 (209 amino acids) makes it most analogous to the rat and human GST P or π subunits (210 amino acids). These proteins, however, do not share much amino acid sequence homology. Indeed, *Drosophila* GST subunit 1 does not share any extended amino acid sequence homology to the mammalian or parasitic GST sequences in the literature, except for 8 (Arg-14, Pro-54, Asp-58, Ala-68, Ile-69, Gly-144, Asp-151, and Leu-160 in Fig. 1) of the 12 amino acid residues conserved in all GST sequences known to date (2–6). Interestingly, subunit 1 does have significant homology to the maize GST III sequence, as shown in Fig. 4 (35). In a region of 44 amino acids, 22 amino acids were identical between the two GSTs at homologous positions (50% identity). In addition, there are 7 amino acids with similar side-chain groups and 6 more with similar polarity for a maximum of 79% sequence homology. This conservation is striking considering the evolutionary difference between maize and *Drosophila melanogaster*. It is therefore tempting to propose functional significance for this stretch of 44 amino acids.

The major difference in chromatographic behavior between maize GST III and *Drosophila* GST 1-1 is that maize GST III can bind to a GSH–agarose affinity column but not to a *S*-hexyl-GSH–agarose affinity column (ref. 36; K. P. Timmerman and C-P.D.T., unpublished results). Neither enzyme has GSH peroxidase activity. Therefore, these 44 conserved residues may be important in binding xenobiotic substrates. On the other hand, the difference in chromatographic behavior may reflect a subtle difference between the two GSTs in binding the common substrate GSH. The evolutionary significance of this sequence conservation remains to be elucidated.

The pGTDm1 cDNA has apparently detected a multigene family on genomic Southern blots. We have also isolated two different, but potentially overlapping, genomic clones (Fig. 3). Partial sequence analysis of the genomic clone λGTDm102 has revealed two sequences homologous to the N-terminal region of pGTDm1 cDNA sequence (C-P.D.T., N. M. Simkovich, Y-P.S.T., and L. Tu, unpublished results). The lack of immunological cross-reactivity between antibody against peak I GST and other potential GST subunits suggests that the 28.5-kDa band, if indeed a GST subunit, may not be closely related to GST subunit 1. Therefore, *Drosophila* GSTs may be encoded by different gene families, which likely

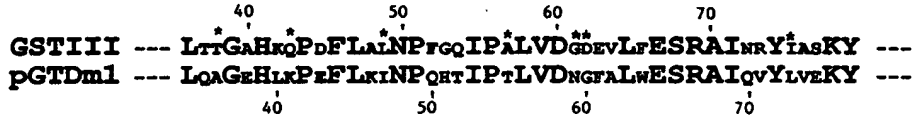



FIG. 4. Partial amino acid sequence comparison between maize GST III and *Drosophila* GST 1–1. The two sequences (in one-letter code) are aligned for maximal homology. Identical amino acids are represented by boldfaced letters. Those amino acid residues that are in the same group of side-chain type (i.e., small polar, S, G, D, and N; large polar, E, Q, K, and R; intermediate polarity, Y, H, and W; large nonpolar, F, M, L, I, and V; and small nonpolar, C, P, A, and T) are labeled with asterisks.

constitute a gene superfamily, as is the case for mammalian systems. GSH *S*-transferase(s) has been suggested to impart pesticide resistance against dichlorodiphenyltrichloroethane (DDT) and organophosphorus compounds in houseflies (*Musca domestica* L.). Molecular analysis of *Drosophila* GSTs should provide further understanding of a family of important xenobiotic metabolizing enzymes and their relationship to pesticide resistance.

We thank Dr. P. Salvaterra for providing the λgtll cDNA library, Loretta Tu for skillful technical assistance, Dr. N. N. Aronson for critical reading of the manuscript, and Eileen McConnell for typing the manuscript. This work was supported by grants from the U.S. Public Health Services [ES02678, ES00140 (RCDA), and USDA 86-CRCR-1-2148 to C-P.D.T. amd GM29006 to T-s.H.]. Part of this work was carried out in 1987 when T-s.H. and C-P.D.T. were in residence at the Institute of Molecular Biology, Academia Sinica in Taipei. The support of Academia Sinica is also gratefully acknowledged.

1. Mannervik, B. & Danielson, U. H. (1988) *CRC Crit. Rev. Biochem.* **23**, 281–334.
2. Pickett, C. B. & Lu, A. Y. H. (1989) *Annu. Rev. Biochem.* **58**, 743–764.
3. DeJong, J. L., Chang, C.-M., Whang-Peng, J., Knutsen, T. & Tu, C.-P. D. (1988) *Nucleic Acids Res.* **16**, 8541–8554.
4. DeJong, J. L., Morgenstern, R., Jornvall, H., DePierre, J. W. & Tu, C.-P. D. (1988) *J. Biol. Chem.* **263**, 8430–8436.
5. Lai, H.-C. J., & Tu, C.-P. D. (1986) *J. Biol. Chem.* **261**, 13793–13799.
6. Lai, H.-C. J., Grove, G. & Tu, C.-P. D. (1986) *Nucleic Acids Res.* **14**, 6101–6114.
7. Frear, D. S. & Swanson, H. R. (1970) *Phytochemistry* **9**, 2123–2132.
8. Shimabukuro, R. H., Swanson, H. R. & Walsh, W. C. (1970) *Plant Physiol.* **46**, 103–107.
9. Shimabukuro, R. H., Frear, D. S., Swanson, H. R. & Walsh, W. C. (1971) *Plant Physiol.* **47**, 10–14.
10. Frear, D. S. & Swanson, H. R. (1973) *Pestic. Biochem. Physiol.* **3**, 473–482.
11. Clark, A. G. & Dauterman, W. C. (1982) *Pestic. Biochem. Physiol.* **17**, 307–314.
12. Hayaoka, T. & Dauterman, W. C. (1983) *Pestic. Biochem. Physiol.* **19**, 344–354.
13. Smith, D. B., Davern, K. M., Board, P. G., Tiu, W. V., Garcia, E. G. & Mitchell, G. F. (1986) *Proc. Natl. Acad. Sci. USA* **83**, 8703–8707.
14. Balloul, J.-M., Sondermeyer, P., Dreyer, D., Capron, M., Grzych, J. M., Pierce, R. J., Carvallo, D., Leconcq, J. P. & Capron, A. (1987) *Nature (London)* **326**, 149–153.
15. Taylor, J. B., Vidal, A., Torpier, G., Meyer, D. J., Roitsch, C., Balloul, J.-M., Southan, C., Sondermeyer, P., Pemble, S., Lecocq, L. P., Capron, A. & Ketterer, B. (1988) *EMBO J.* **7**, 465–472.
16. Wistow, G. & Piatigorsky, J. (1987) *Science* **236**, 1554–1556.
17. Tomarev, S. I. & Zinovieva, R. D. (1988) *Nature (London)* **336**, 86–88.
18. Eschalier, G. & Ohanessian, A. (1970) *In Vitro* **6**, 162–172.
19. Itoh, N., Salvaterra, P. & Itakura, K. (1985) *Drosophila Inf. Serv.* **61**, 89.
20. Maniatis, T., Fritsch, E. F. & Sambrook, J. (1982) *Molecular Cloning: A Laboratory Manual* (Cold Spring Harbor Lab., Cold Spring Harbor, NY).
21. Tu, C.-P. D. & Reddy, C. C. (1985) *J. Biol. Chem.* **260**, 9961–9964.
22. Reddy, C. C., Burgess, J. R., Gong, Z. Z., Massaro, E. J. & Tu, C.-P. D. (1983) *Arch. Biochem. Biophys.* **224**, 87–101.
23. Tu, C.-P. D., Weiss, M. J., Li, N. & Reddy, C. C. (1983) *J. Biol. Chem.* **258**, 4659–4662.
24. Towbin, H., Staehelin, T. & Gordon, J. (1979) *Proc. Natl. Acad. Sci. USA* **76**, 4350–4354.
25. Tu, C.-P. D., Matsushima, A., Li, N., Rhoads, D. M., Srikumar, K., Reddy, A. P. & Reddy, C. C. (1986) *J. Biol. Chem.* **261**, 9540–9545.
26. Young, R. A. & Davis, R. W. (1983) *Science* **222**, 778–782.
27. Blin, N., Gabain, A. V. & Bujard, H. (1975) *FEBS Lett.* **53**, 84–86.
28. Southern, E. M. (1975) *J. Mol. Biol.* **98**, 503–517.
29. Sanger, F., Nicklen, S. & Coulson, A. R. (1977) *Proc. Natl. Acad. Sci. USA* **74**, 5463–5467.
30. Messing, J. (1983) *Methods Enzymol.* **100**, 20–79.
31. Biggin, M. D., Gibson, T. J. & Hong, G. F. (1983) *Proc. Natl. Acad. Sci. USA* **80**, 3963–3965.
32. Lai, H.-C. J., Qian, B., Grove, G. & Tu, C.-P. D. (1988) *J. Biol. Chem.* **263**, 11389–11395.
33. Chow, N.-W. I., Whang-Peng, J., Kao-Shan, C.-S., Tam, M. F., Lai, H.-C. J. & Tu, C.-P. D. (1988) *J. Biol. Chem.* **263**, 12797–12800.
34. Cochrane, B. J., Morrissey, J. J. & LeBlanc, G. A. (1987) *Insect Biochem.* **17**, 731–738.
35. Grove, G., Zarlengo, R. P., Timmerman, K. P., Li, N., Tam, M. F. & Tu, C.-P. D. (1988) *Nucleic Acids Res.* **16**, 425–438.
36. Mozer, T. J., Tiemeier, D. C. & Jaworski, E. G. (1983) *Biochemistry* **22**, 1068–1072.

*Proc. Natl. Acad. Sci. USA*
Vol. 90, pp. 4404–4408, May 1993
Biochemistry

# The LIM motif defines a specific zinc-binding protein domain

(metalloprotein/cysteine-rich protein motifs/cytoskeletal protein)

JAMES W. MICHELSEN*, KAREN L. SCHMEICHEL*, MARY C. BECKERLE*†, AND DENNIS R. WINGE†‡

Departments of *Biology, and of ‡Medicine and Biochemistry, University of Utah, Salt Lake City, UT 84132

*Communicated by Sidney Velick, February 3, 1993*

ABSTRACT **The cysteine-rich protein (CRP) contains two copies of the LIM sequence motif, $CX_2CX_{17}HX_2CX_2CX_2CX_{17}CX_2C$, that was first identified in the homeodomain proteins Lin-11, Isl-1, and Mec-3. The abundance and spacing of the cysteine residues in the LIM motif are reminiscent of a metal-binding domain. We examined the metal-binding properties of CRP isolated from chicken smooth muscle (cCRP) and from a bacterial expression system and observed that cCRP is a specific Zn-binding metalloprotein. Four Zn(II) ions are maximally bound to cCRP, consistent with the idea that each LIM domain coordinates two metal ions. From spectroscopic studies of Co(II)- and $^{113}$Cd(II)-substituted cCRP, we determined that each metal ion is tetrahedrally coordinated with cysteinyl sulfurs dominating the ligand types. One metal site within each LIM motif has tetrathiolate ($S_4$) coordination, the second site may either be $S_4$ or $S_3N_1$. The LIM motif represents another example of a specific Zn-binding protein sequence.**

The LIM motif is a cysteine-rich sequence found in a diverse collection of proteins including transcription factors (1–5), a protooncogene product (6, 7), and cytoskeletal components (8–11). Many of the LIM proteins appear to be involved in regulation of gene expression and cellular differentiation during development. The specific function of the LIM domain has not been established, although it has been postulated to serve as a DNA or protein binding interface. Because of the abundance of conserved cysteine residues in the LIM consensus sequence, the motif has been widely proposed to be a metal-binding sequence. Efforts have been made to examine the metal-binding properties of LIM-motif proteins (12, 13). For example, it has been demonstrated that the cysteine-rich intestinal protein (CRIP) will bind exogenously applied Zn ions (13). In addition, expression of the LIM domains of Lin-11 in *Escherichia coli* resulted in the isolation of a protein from inclusion bodies that contained both Zn and an Fe-S cluster (12). The observation of an Fe-S cluster in Lin-11 prompted consideration of the intriguing idea that the LIM transcription factors might be redox-regulated (12).

In this report we present a comprehensive metal analysis of a LIM-domain protein isolated from its endogenous source. Specifically, we describe the metal-binding properties of the chicken cysteine-rich protein (cCRP) isolated from avian smooth muscle (10). cCRP, the chicken homologue of the human CRP (11), exhibits two LIM domains of the sequence $CX_2CX_{17}HX_2CX_2CX_2CX_{17}CX_2C$ (J. Pino and M.C.B., unpublished results). Our results show cCRP to be a Zn(II) metalloprotein. The implications of these results for the LIM domain structure are discussed.

## MATERIALS AND METHODS

**Purification of cCRP.** cCRP was purified from fresh chicken gizzards by a procedure to be described in detail elsewhere (A. W. Crawford, J. D. Pino, and M.C.B., unpublished work). The purity of cCRP was demonstrated by SDS/PAGE and amino acid analysis. An extinction coefficient for cCRP of $2.66 \times 10^4$ $M^{-1}\cdot cm^{-1}$ was obtained by measurement of the absorbance at 280 nm followed by quantitation of the cCRP protein by amino acid analysis. Thiol titrations were carried out as described (14, 15).

**Metal Exchange.** cCRP was prepared for metal exchange reactions by dialysis in buffer M (40 mM Tris Cl, pH 7.5/40 mM KCl). The protein was diluted 10-fold with 0.2 M potassium phosphate (pH 7.2) and subsequently incubated with mentioned quantities of metal salts. Spectra were recorded after 10 min. For measurements of binding stoichiometry, the metal-replaced samples were incubated for 1–3 hr, dialyzed in buffer M containing 0.1 mM EDTA, and subsequently quantified for metals and protein.

Cd(II), Zn(II), and Co(II) stock solutions were made in 1 mM HCl, and concentrations were verified by atomic absorption spectroscopy. $^{113}$Cd was obtained as the oxide and was converted to the chloride salt before use. Cu(I) stock solutions were prepared in 2.5 mM HCl/0.2% NaCl.

**Metal Reconstitution.** cCRP was denatured in 6 M guanidinium chloride and, after gel filtration on Sephadex G-50, the demetallized and denatured cCRP was reduced with 5 mM dithiothreitol for 12–18 hr at 4°C and subsequently rechromatographed on Sephadex G-50 equilibrated in 6 M guanidinium chloride at pH 5.0. Metal was added to the denatured protein to the desired equivalency. The samples, incubated for 30 min at 25°C, were subsequently diluted 10-fold with the addition of 0.2 M potassium phosphate (pH 7.2) or 0.2 M Tris Cl (pH 9). Reconstitutions with Cd(II) or Zn(II) were performed under aerobic conditions with the phosphate buffer, whereas Co(II) and Cu(I) reconstitutions were carried out anaerobically with the Tris buffer.

**Spectroscopy.** Ultraviolet spectroscopy was carried out on a Beckman DU spectrometer. Luminescence measurements were made on a Perkin–Elmer 650-10S fluorimeter. $^{113}$Cd NMR spectroscopy was performed on a Unity 500 Varian spectrometer operating in the Fourier transform mode at 110.9 MHz. Spectra were recorded on $^{113}$Cd samples (8–12 mg/ml) containing 50% $^2H_2O$ as a field lock.

**Bacterial Expression of cCRP.** A cCRP cDNA clone that was complete with respect to coding capacity was obtained by screening a chicken embryo fibroblast cDNA library (16) with a human CRP cDNA probe (J. D. Pino and M.C.B., unpublished results). The cCRP coding sequence was amplified by the polymerase chain reaction (PCR) and cloned into pET-5b (17). The authenticity of the expression construct, termed pET-cCRP, was confirmed by restriction mapping, PCR, and DNA sequencing.

For expression of cCRP sequences in bacteria, pET-cCRP was transformed into *E. coli* BL21(DE3)/pLysS. After in-

The publication costs of this article were defrayed in part by page charge payment. This article must therefore be hereby marked "*advertisement*" in accordance with 18 U.S.C. §1734 solely to indicate this fact.

Abbreviations: CRP, cysteine-rich protein; cCRP, chicken CRP; LMCT, ligand → metal charge transfer; IPTG, isopropyl β-D-thiogalactopyranoside.

†To whom reprint requests should be addressed.

duction of cCRP expression with isopropyl β-D-thiogalactopyranoside (IPTG), the cells were collected by centrifugation, were washed, and were lysed in 25 mM Tris Cl, pH 7.5/150 mM NaCl/10 mM dithiothreitol by a rapid freeze–thaw cycle followed by sonication. The lysates were centrifuged for 30 min at 13,000 × *g* to separate soluble components (supernatant) from insoluble pelletable material. cCRP was found exclusively in the soluble fraction after centrifugation. The soluble material was dialyzed against buffer B-10 (20 mM Tris acetate, pH 7.6/20 mM NaCl/0.1 mM EDTA/0.1% 2-mercaptoethanol) and cCRP was isolated by ion-exchange chromatography on DEAE-cellulose. Fractions containing cCRP were pooled and dialyzed against buffer B-10. The metal content of the protein was determined as described above. The metal content of the pellets derived from the cell lysates was also examined. The brown-red pellets were washed twice with brief sonication in lysis buffer containing 4 M urea. The residual, pelletable material was solubilized with 7 M urea in preparation for metal analysis.

## RESULTS

**Metal Content of cCRP.** cCRP was purified from avian smooth muscle as a monomer with an apparent molecular mass of 23 kDa (Fig. 1, lane 1). By atomic absorption spectroscopy, Zn(II) was observed to be specifically associated with cCRP. The average Zn(II) content was found to be 2.9 ± 0.45 mol per mol of cCRP (Table 1). All samples of cCRP examined contained bound Zn(II); however, some variability in the Zn(II) stoichiometry was observed as a result of protein oxidation that occurred during the isolation procedure. A correlation was observed between the number of titratable thiols and the observed Zn(II) binding stoichiometry. For example, Zn-cCRP isolates with 3.4 mol eq. of Zn(II) contained 11.6 titratable thiols, whereas isolates with Zn(II) stoichiometries of 2.7 and 2.4 contained 9.5 and 8.5 thiol groups, respectively. The bound Zn(II) content of native cCRP did not change with the addition of 5 mol eq. of Zn(II), but preincubation of Zn-cCRP with 5 mM dithiothreitol followed by addition of Zn(II) increased the Zn(II) content of cCRP to 4 mol eq. (Table 1). The Zn(II) ions were not depleted by prolonged dialysis in 0.1 mM EDTA at pH > 6.8, which indicated that the Zn(II) ions were tightly bound. In contrast with the report that LIM domains derived from

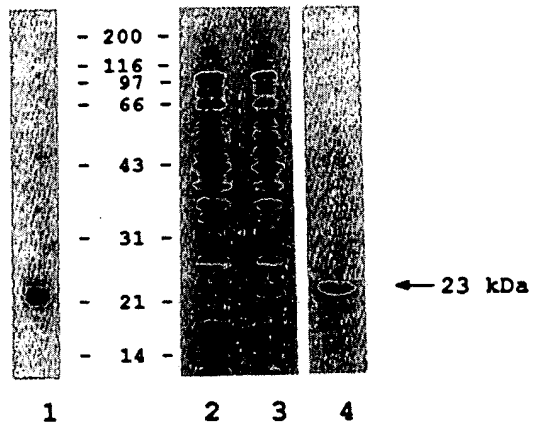



FIG. 1. Purification of Zn-cCRP from avian smooth muscle and the bacterial expression system. Proteins were separated by SDS/polyacrylamide gel electrophoresis and detected by staining with Coomassie blue. Lane 1, cCRP purified from smooth muscle; lane 2, bacterial proteins expressed in the absence of IPTG; lane 3, bacterial proteins expressed upon IPTG induction; lane 4, cCRP purified from the bacterial expression system.

Table 1. Metal stoichiometry of cCRP samples

| Sample | mol eq. added | Final stoichiometry, mol/mol Zn | Cd | *n** |
|---|---|---|---|---|
| Native | — | 2.9 ± 0.45 | 0 | 10 |
| $Zn_{2.4}$-cCRP | 5 Zn(II) | 2.4 | 0 | 1 |
| $Zn_{2.7}$-cCRP (+ DTT) | 5 Zn(II) | 4.2 | 0 | 1 |
| Apo (+ DTT) | 10 Zn(II) | 4.3 ± 0.03 | 0.1 ± 0.1 | 2 |
| Apo (+ DTT) | 10 Cd(II) | 0.3 ± 0.03 | 3.6 ± 0.2 | 2 |
| Bacterially expressed cCRP | — | 3.8 ± 0.1 | | 3 |

The metal ion stoichiometry of Zn-cCRP (mean ± SD) was determined in the presence and absence of added metal salts. In experiments with added metal, the number specified refers to the mol eq. of added metal and the particular metal ion added. The final stoichiometry numbers refer to values obtained after dialysis in 0.1 mM EDTA buffered at pH 7.5. DTT, dithiothreitol.
*No. of measurements.

Lin-11 exhibit both Zn and Fe (12), we observed no Fe associated with cCRP.

To evaluate whether the heterologous expression of Lin-11 in bacteria may have resulted in the Fe binding observed by Li *et al.* (12), we examined the metal content of cCRP isolated from a bacterial expression system (Fig. 1, lanes 2–4). Analysis of the metal content of bacterially expressed cCRP revealed that 3.8 ± 0.1 mol of Zn(II) were present per mol of protein (Table 1). As was the case for cCRP purified from smooth muscle, no Fe or Cu was detected in any isolate. The addition of ferric ammonium citrate at 0.6 mM to the growth medium did not result in any Fe in cCRP. Thus, the LIM-domain protein cCRP is a specific Zn(II) metalloprotein whether it is purified from its endogenous source or from a heterologous expression system.

**Biophysical Studies of cCRP.** Biophysical studies were carried out on smooth muscle cCRP to probe the LIM metal centers. Titrations of a Zn-cCRP sample containing 3.4 mol eq. of bound Zn(II) with 10 mol eq. of added Cd(II) led to a facile metal exchange resulting in 3.2 mol eq. of Cd(II) bound after dialysis in 0.1 mM EDTA. A 1:1 molar displacement of bound Zn(II) ions was observed. Cd-cCRP molecules exhibited transitions in the ultraviolet consistent with cysteine-Cd(II) ligand → metal charge transfer (LMCT) bands. Zn(II) displacement was also observed by titrations with Cu(I) ions at neutral pH. Cu-cCRP formed by the addition of Cu(I) exhibited ultraviolet transitions characteristic of S → Cu(I) LMCT bands and luminescence with similar emission properties to proteins with Cu(I)-thiolate complexes (18).

cCRP can be fully depleted of bound Zn(II) ions by dialysis at pH 4 or by denaturation in 6 M guanidinium chloride. Denatured, apo-cCRP could be repopulated with metal upon reduction followed by rapid dilution into concentrated buffer containing metal ions. Approximately 4 mol eq. of Zn(II) or Cd(II) were bound after reconstitution and dialysis in 0.1 mM EDTA (Table 1).

Denatured apo-cCRP was reconstituted with Co(II) to probe the coordination geometry of the binding sites (Fig. 2). Electronic spectroscopy of Co(II) complexes is useful in elucidating the coordination number and geometry around the metal ion (19). Apo-cCRP reconstituted anaerobically with increasing quantities of Co(II) at pH 8.5 yielded blue-green protein samples. The electronic spectrum of $Co_2$-cCRP was dominated by *d–d* transitions with maxima at 620 nm [$\epsilon$ = 513 $M^{-1}\cdot cm^{-1}$ per Co(II)], 701 nm ($\epsilon$ = 662), and 740 nm ($\epsilon$ = 621) (Fig. 2*A*). The energy of these transitions and the molar extinction coefficients are typical of the spin-allowed $\upsilon_3[^4A_2 \rightarrow {}^4T_1(P)]$ transitions of four coordinate Co(II) complexes in distorted tetrahedral geometry (19–24). The presence of the prominent 740-nm component of the *d–d* band is

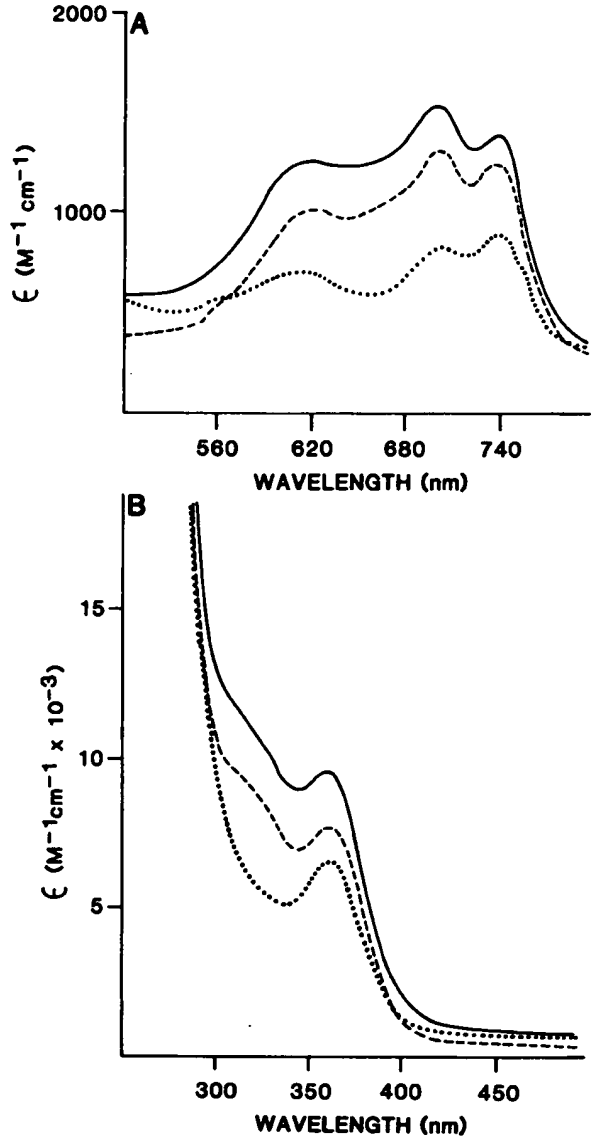



FIG. 2. Electronic spectra of Co(II)-cCRP. Co(II) was added to apo-cCRP in 6 M guanidinium chloride at pH 5. The mixture was diluted 20-fold in buffer M at pH 9 and spectra were measured in anaerobic cuvettes. The final protein concentration was 53 $\mu$M. The curves in *A* and *B* refer to Co-cCRP samples at 1 (· · ·), 2 (- - -), and 4 (—) mol eq. of Co(II). The extinction coefficient plotted on the ordinate is based on the cCRP concentration. *A* shows the *d–d* spectral region, and *B* shows the charge-transfer spectral region.

consistent with predominantly thiolate coordination (19–25). The intensity of the *d–d* absorbance envelope was enhanced with increasing amounts of added Co(II) from 1 to 4 mol eq. (Fig. 2*A*). Co(II)-thiolate ligation is also suggested by features in the near ultraviolet spectral region. Charge transfer transitions in Co(II)-thiolate clusters usually occur between 340 and 360 nm with molar extinction coefficients of 800–1300 $M^{-1}$·$cm^{-1}$ per Co-S-'Cys' bond (20–24). A prominent peak was observed in Co-cCRP samples near 360 nm (Fig. 2*B*). The extinction coefficients per Co-S-Cys bond for $Co_1$-cCRP and $Co_2$-cCRP were 1636 and 970, respectively (Fig. 2*B*). The intensity and position of this transition are consistent with LMCT charge transfer bands of Co(II)-thiolate complexes (25). The prominent LMCT band and the 740-nm component of the *d–d* envelope in the $Co_1$-cCRP sample suggest that the initial Co(II) ion is ligated by four thiolates.

Additional information on the metal-binding properties of cCRP was obtained by $^{113}$Cd NMR. The chemical shift of a $^{113}$Cd(II) signal is sensitive to the structure and ligand type of the metal binding site (26). $^{113}$Cd NMR is particularly useful in proteins with cysteinyl thiolate ligands, as thiolates induce large downfield shifts in the $^{113}$Cd resonances (26). In cCRP, each LIM sequence contains eight conserved residues capable of metal ion ligation, seven of which are cysteines. Spectra were obtained on $^{113}$Cd-cCRP samples prepared by metal exchange and reconstitution from apo-cCRP. Analysis of a $^{113}Cd_2Zn_{1.8}$-cCRP sample prepared by metal exchange revealed four $^{113}$Cd(II) signals with chemical shifts of 707, 705, 656, and 647 ppm relative to $Cd(ClO_4)_2$ (Fig. 3, spectrum A). The downfield shift of the signals at 707 and 705 ppm is consistent with tetrathiolate Cd(II) coordination (26–32). $^{113}$Cd(II) resonances with tetrathiolate ligation occur at 704 and 710 ppm in the glucocorticoid receptor, and at 707 and 659 ppm in the yeast transcription factor GAL4 (27, 29). The $^{113}$Cd resonances at 656 and 647 ppm could arise from either tetrathiolate ($S_4$) coordination or $S_3N$ coordination (26–32). Cd(II) coordination with an $S_3N$ ligand field is seen in gene 32 protein (637 ppm $^{113}$Cd resonance) of bacteriophage T4 and in the nucleocapsid protein ($^{113}$Cd chemical shifts of 659 and 640 ppm) of human immunodeficiency virus (31, 32). $Cd_2$-cCRP prepared with $^{113}$Cd(II) by reconstitution of the denatured, reduced apo-cCRP revealed the same four $^{113}$Cd resonances as in the sample prepared by metal displacement (Fig. 3, spectrum B). A $Cd_4$-cCRP sample prepared by reconstitution from the denatured sample revealed only two signals, at 707 and 646 ppm (Fig. 3, spectrum C). These two chemical shifts were similar to two of the four chemical shifts observed in $Cd_2$-cCRP samples.

## DISCUSSION

The LIM-motif protein cCRP was shown to be a Zn(II)-containing metalloprotein. Zn was the only metal ion associated with cCRP in isolates from avian smooth muscle and in isolates from bacteria engineered to express the cCRP gene. The maximal Zn(II) stoichiometry is likely to be 4 mol eq. This conclusion is based on two observations. First, the metal content of cCRP after pre-reduction and addition of excess metal ion is near 4 mol eq. Second, the Zn(II) content in the bacterially expressed cCRP was 3.8 mol eq. The Zn(II) content of the bacterially expressed cCRP may be higher than that in avian muscle isolates due to the rapid purification of Zn-cCRP from the bacterial expression system, a feature that could minimize the extent of protein oxidation.

The Zn(II) ions are ligated to cCRP predominantly through cysteinyl thiolates. This conclusion is based on the characteristic S→Cd(II) and S→Co(II) charge transfer bands in the ultraviolet and near ultraviolet, respectively, and the chemical-shift range of $^{113}$Cd(II) nuclei substituted in cCRP. The *d–d* band region of Co(II)-cCRP and chemical shift of $^{113}$Cd(II) signals indicate that the metal ions are bound in a tetrahedral coordination geometry.

cCRP contains two LIM motifs and maximally binds 4 mol eq. of Zn(II) or Cd(II), suggesting that each LIM motif may accommodate up to two Zn(II) ions. It is curious that cCRP samples saturated with Cd(II) yield only two $^{113}$Cd resonances whereas samples with subsaturating amounts of Cd(II) yield four $^{113}$Cd resonances. It is possible that the $^{113}$Cd NMR spectrum of $Cd_4$-cCRP reflects molecular symmetry with the two LIM domains being independent and approximately equivalent with respect to the metal coordination geometry and environment. According to this interpretation, the resonances at 707 and 646 ppm would occur when $^{113}$Cd(II) ions are present at both metal sites within a single LIM domain. The two LIM domains in cCRP must then adopt very similar tertiary folds. Four resonances in

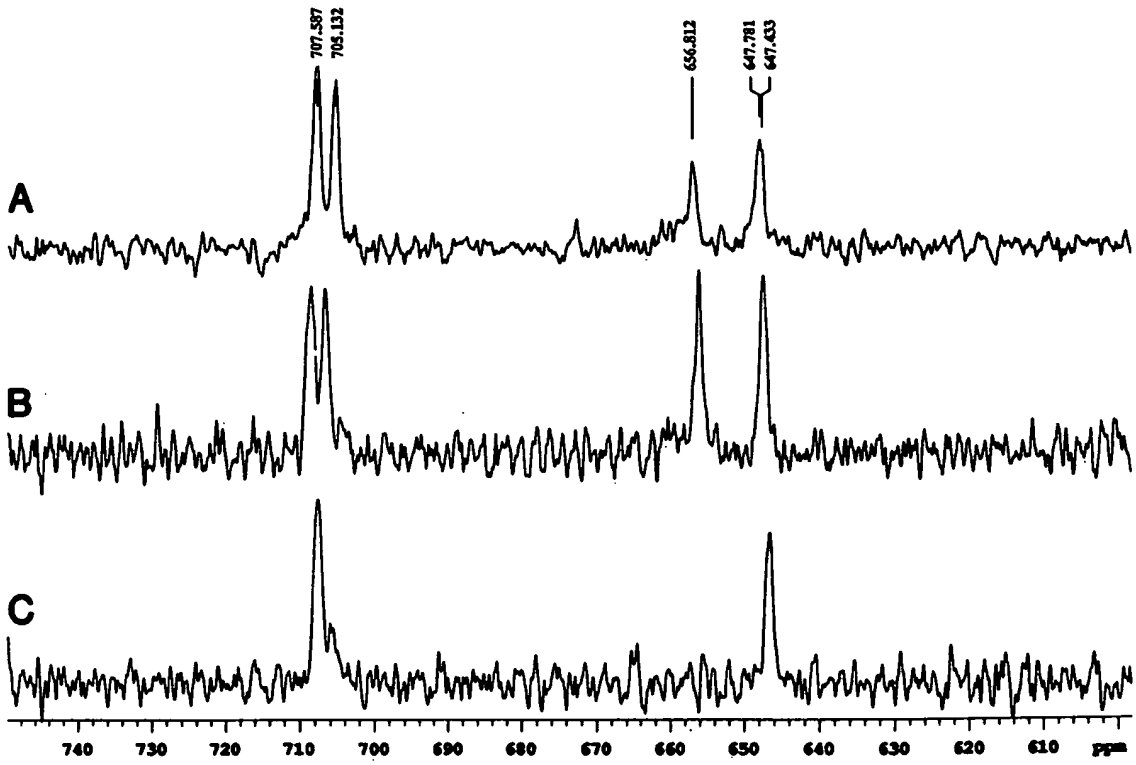



FIG. 3. NMR spectroscopy of $^{113}$Cd-cCRP samples at pH 7.5 in buffer M. Spectrum A is of $^{113}Cd_2Zn_{1.8}$-cCRP (0.8 mM cCRP) prepared by metal exchange, spectrum B is of $^{113}Cd_{1.8}$-cCRP (0.7 mM) prepared by reconstitution of apo-cCRP, and spectrum C is of $^{113}Cd_{4.2}$-cCRP (0.8 mM). Chemical shifts are relative to 1 M $Cd(ClO_4)_2$. The number of transients was 90,000 for each spectrum. Samples contained 50% $^2H_2O$ as a field lock. Parameters used include a spectral width of 33.2 kHz (300 ppm) for the first spectrum and 16.8 kHz (150 ppm) for subsequent spectra, an acquisition time of 0.8 sec, and a 60° pulse width. A 40-Hz line broadening was applied for spectral enhancement. Increasing the acquisition time to 1.5 sec did not appreciably affect the signal-to-noise ratio, implying that the $T_1$ is <1 sec.

subsaturated molecules may reflect a population of molecules differing in the sites occupied by $^{113}$Cd(II) ions. The 707- and 646-ppm resonances may arise from occupancy of $^{113}$Cd(II) ions at each of the two LIM domain sites. The 705- and 656-ppm resonances may arise from a $^{113}$Cd(II) ion occupying only one of the two sites, with the second site either vacant or occupied by Zn(II). There may exist a slight preference energetically for binding of two Cd(II) ions within the same LIM domain.

The amino acid sequences of cCRP's two LIM domains are shown in Fig. 4. All LIM domains analyzed to date exhibit eight spatially aligned potential metal ligands, with an absolutely conserved histidine in consensus position 3. We postulate that each LIM domain of cCRP comprises a $Cys_3His_1$ Zn(II) site and a $Cys_4$ Zn(II) site. The eight LIM consensus residues are theoretically sufficient to bind the Zn(II), although each LIM domain exhibits additional potential metal ligands. While we cannot unequivocally assign the ligands, the consensus residues are likely to be prominently involved in metal binding.

The exclusive presence of Zn(II) in cCRP contrasts with the observation of an Fe-S cluster in bacterially expressed Lin-11 (12). It has not been determined whether Lin-11 isolated from *Caenorhabditis elegans* contains an Fe-S cluster or whether, like cCRP, it is a specific Zn-binding metalloprotein *in vivo*. It is possible that sequence differences in the LIM domains derived from Lin-11 and cCRP define differences in the metal-binding capacity and specificity of the two proteins. It is also possible that the association of Fe with Lin-11 may be an artifact of heterologous expression. Misinsertion of metal ions into metalloproteins synthesized in a heterologous system has been documented. For example, isolation of azurin expressed in bacteria yielded a Zn metalloprotein rather than a native Cu metalloprotein (33). Lin-11

```
LIM-1:     CGVCQKAVYFAEEVQCEGSSFHKSCFLCMVCKKNLDSTTVAVHGDEIYCKSC

LIM-2:     CPRCGQAVYAAEKVIGAGKSWHKSCFRCAKCGKSLESTTLADKDGEIYCKGC

  CRP
Consensus: CX2C.....X17.....HX2CX2CX2C.....X17.....CX2C

  LIM
Consensus: CX2C.....X16-23.....HX2CX2CX2C.....X16-21.....CX2-3(C,H,D)

Consensus
 Position: 1..2.....3..4..5..6.....7..8
```

FIG. 4. The LIM domains of cCRP. The amino acid sequences of both LIM domains of cCRP are shown, with LIM-1 representing the more N-terminal LIM motif. Eight residues that are found in all LIM domains to date are outlined and are numbered based on their position in the linear sequence. Other potential metal-liganding residues in cCRP are underlined. The cCRP LIM consensus is compared with the LIM consensus derived from all known LIM sequences.

LIM domains may also have bound Fe during their isolation from the bacterial cell lysate. Unlike cCRP, which is a soluble protein when expressed in bacteria, bacterially expressed Lin-11 was sequestered into inclusion bodies necessitating a denaturation–renaturation step prior to metal analysis. We observed that pellets derived from bacterial lysates contain significant amounts of Fe. Moreover, the urea procedure described by Li *et al.* (12) for isolation of Lin-11 results in the solubilization of Fe that can be partially depleted with Chelex 100. Therefore it is possible that Lin-11 acquired Fe during solubilization. In any event, our findings demonstrate unequivocally that not all proteins displaying LIM domains will be regulated by cellular redox conditions as was proposed for Lin-11.

Metal ions are known to stabilize tertiary conformations of proteins, participate in catalysis in certain metalloenzymes, and induce changes in protein structure. A variety of Zn-finger structures, including the binuclear Zn(II) sites in GAL4 and the Zn(II) center in the steroid hormone receptor domain, are dependent on Zn(II) to stabilize their tertiary folds (34–37). Zn(II) binding may likewise play an essential role in stabilizing the tertiary structure of LIM domains within proteins to generate a binding interface for proteins or nucleic acids.

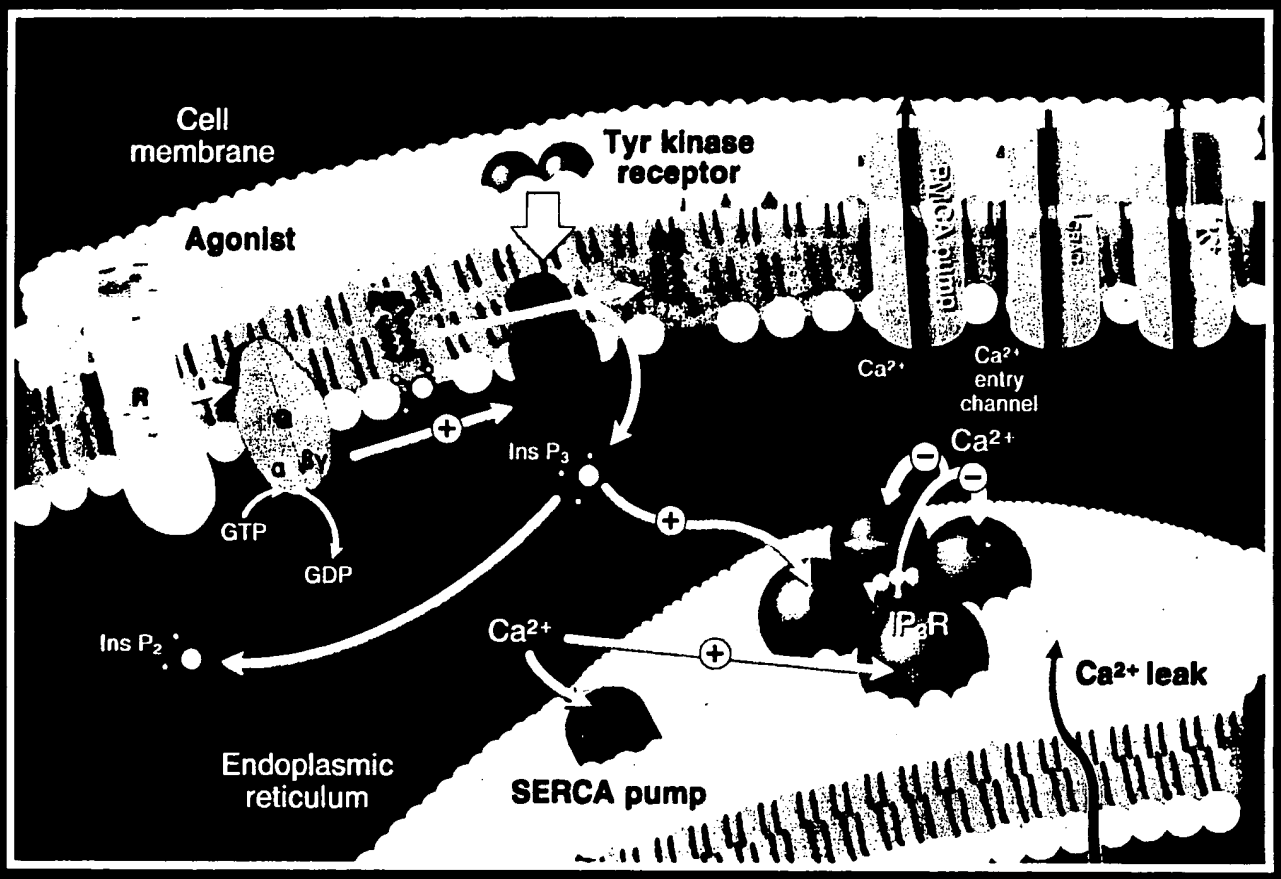
We thank J. Olson for NMR technical assistance; Drs. I. M. Armitage, D. Davis, and J. Otvos for discussions on the $^{113}Cd$ NMR experiments; Dr. J. Tamkun for the cDNA library; and Drs. N. Cooke and S. Liebhaber for human CRP cDNA. We are grateful to J. Pino for providing cCRP cDNA and allowing us to cite unpublished sequence information. This work was supported by funds from the American Heart Association (to M.C.B.), the National Institutes of Health (HL41553 to M.C.B. and ES03817 to D.R.W.), the Utah Regional Cancer Center (CA42014), and a predoctoral cancer training grant (5T32 CA 09602 to K.L.S.). M.C.B. is an Established Investigator of the American Heart Association.

1. Freyd, G., Kim, S. K. & Horvitz, R. (1990) *Nature (London)* **344**, 876–879.
2. Karlsson, O., Thor, S., Norberg, T., Ohlsson, H. & Edlund, T. (1990) *Nature (London)* **344**, 879–882.
3. Way, J. C. & Chalfie, M. (1988) *Cell* **54**, 5–16.
4. Taira, M., Jamrich, M., Good, P. J. & Dawid, I. B. (1992) *Genes Dev.* **6**, 356–366.
5. Cohen, B., McGuffin, E., Pfeifle, C., Segal, D. & Cohen, S. M. (1992) *Genes Dev.* **6**, 715–729.
6. Boehm, T., Foroni, L., Kaneko, Y., Perutz, M. F. & Rabbitts, T. H. (1991) *Proc. Natl. Acad. Sci. USA* **88**, 4367–4371.
7. McGuire, E. A., Rintoul, C. E., Sclar, G. M. & Korsmeyer, S. J. (1992) *Mol. Cell. Biol.* **12**, 4186–4196.
8. Crawford, A. W. & Beckerle, M. (1991) *J. Biol. Chem.* **266**, 5847–5853.
9. Crawford, A. W., Michelsen, J. W. & Beckerle, M. (1992) *J. Cell Biol.* **116**, 1381–1393.
10. Sadler, I., Crawford, A. W., Michelsen, J. W. & Beckerle, M. C. (1992) *J. Cell Biol.* **119**, 1573–1587.
11. Liebhaber, S. A., Emery, J. G., Urbanek, M., Wang, X. & Cooke, N. E. (1990) *Nucleic Acids Res.* **18**, 3871–3879.
12. Li, P. M., Reichert, J., Freyd, G., Horvitz, H. R. & Walsh, C. T. (1991) *Proc. Natl. Acad. Sci. USA* **88**, 9210–9213.
13. Hempe, J. M. & Cousins, R. J. (1991) *Proc. Natl. Acad. Sci. USA* **88**, 9671–9674.
14. Habeeb, A. F. S. A. (1972) *Methods Enzymol.* **25**, 457–464.
15. Grassetti, D. R. & Murray, J. F., Jr. (1967) *J. Biol. Chem.* **119**, 41–49.
16. Tamkum, J. W., DeSimone, D. W., Fonda, D., Patel, R. S., Buck, C., Horwitz, A. F. & Hynes, R. O. (1986) *Cell* **46**, 271–282.
17. Studier, F. W., Rosenberg, A. H., Dunn, J. J. & Dubendoff, J. W. (1990) *Methods Enzymol.* **185**, 60–89.
18. Winge, D. R., Dameron, C. T. & George, G. (1992) *Adv. Inorg. Biochem.* in press.
19. Bertini, I. & Luchinat, C. (1984) *Adv. Inorg. Biochem.* **6**, 72–111.
20. May, S. W. & Juo, J. Y. (1978) *Biochemistry* **17**, 3333–3338.
21. Johnson, R. S. & Schachman, H. K. (1983) *J. Biol. Chem.* **258**, 3528–3538.
22. Giedroc, D. P. & Coleman, J. E. (1986) *Biochemistry* **25**, 4969–4978.
23. Lane, R. W., Ibers, J. A., Frankel, R. B., Papaefthymiou, G. C. & Holm, R. H. (1977) *J. Am. Chem. Soc.* **99**, 84–97.
24. Swenson, D., Baenziger, N. C. & Coucouvanis, D. (1978) *J. Am. Chem. Soc.* **100**, 1932–1934.
25. Dance, I. G. (1979) *J. Am. Chem. Soc.* **101**, 6264–6273.
26. Armitage, I. M., Uiterkamp, A. J. M. S., Chlebowski, J. F. & Coleman, J. E. (1978) *J. Magn. Reson.* **29**, 375–392.
27. Pan, T., Freedman, L. P. & Coleman, J. E. (1990) *Biochemistry* **29**, 9218–9225.
28. Gardner, K. H., Pan, T., Narula, S., Rivera, E. & Coleman, J. E. (1991) *Biochemistry* **30**, 11292–11302.
29. Pan, T. & Coleman, J. E. (1990) *Biochemistry* **29**, 3023–3029.
30. Dalgarno, D. C. & Armitage, I. M. (1984) in *Advances in Inorganic Biochemistry*, eds. Eichhorn, G. & Marzilli, L. G. (Elsevier, New York), Vol. 6, pp. 113–138.
31. Fitzgerald, D. W. & Coleman, J. E. (1990) *Biochemistry* **30**, 5195–5201.
32. Giedroc, D. P., Qiu, H., Khan, R., King, G. C. & Chen, K. (1992) *Biochemistry* **31**, 765–774.
33. Nar, H., Huber, R., Messerschmidt, A., Filippou, A., Barth, M., Jaquinod, M., Kamp, M. & Canters, G. W. (1992) *Eur. J. Biochem.* **205**, 1123–1129.
34. Kraulis, P. J., Raine, A. R. C., Gadhavi, P. L. & Laue, E. D. (1992) *Nature (London)* **356**, 448–453.
35. Marmorstein, R., Carey, M., Ptashne, M. & Harrison, S. C. (1992) *Nature (London)* **356**, 408–414.
36. Lee, M. S., Gippert, G. P., Soman, K. V., Case, D. A. & Wright, P. E. (1989) *Science* **245**, 635–637.
37. Luisi, B. F., Xu, W. X., Otwinowski, Z., Freedman, L. P., Yamamoto, K. R. & Sigler, P. B. (1991) *Nature (London)* **352**, 497–505.

# Cell



Volume 80 Number 2 January 27, 1995




# Signal Transduction Review Issue

# Cell



**Telephone Numbers**

| | |
|---|---|
| **Editorial:** | 617-661-7057 |
| **Advertising:** | 617-661-7059 |
| **Subscriptions:** | 617-661-7060 |

**Principal Office**
Cell
50 Church Street
Cambridge, Massachusetts 02138
USA
Fax: 617-661-7061
E-mail: editor@cell.com

**European Office**
Laboratory of Eukaryotic Molecular Genetics
NIMR
The Ridgeway, Mill Hill
London NW7 1AA, England
Telephone: 81-906-3897
Fax: 81-913-8527
E-mail: p-rigby@nimr.mrc.ac.uk

**Cell** (ISSN 0092-8674) is published biweekly by Cell Press, 50 Church Street, Cambridge, Massachusetts 02138. The institutional subscription rate for 1995 is $375 (US), $525 (Europe and Israel), or $460 (elsewhere). The individual subscription rate is $105 (US), $225 (Europe and Israel), or $170 (elsewhere). The individual copy price is $30.

 Second class postage paid at Boston, Massachusetts, and additional mailing offices. Postmaster: send address changes to Cell, 50 Church Street, Cambridge, Massachusetts 02138.

Volume 80 Number 2 January 27, 1995

# Cell





(*continued*)

Cell, Vol. 80, 237–248, January 27, 1995, Copyright © 1995 by Cell Press

# Modular Binding Domains in Signal Transduction Proteins

Review

**George B. Cohen, Ruibao Ren,* and David Baltimore**
Department of Biology
Massachusetts Institute of Technology
77 Massachusetts Avenue
Cambridge, Massachusetts 02139

The transduction of a signal is a change in form of the signal as it is passed from one carrier to another. The root "duce" means "to lead" in Latin; thus, a signal is led through a cell by steps of transduction (the same root is in the words seduce and duct as well as Il Duce). The earliest transduction steps that were elucidated involved massive release of small molecule "second messengers", originally cAMP, that flooded a cell with information. With the understanding that such proteins as tyrosine kinases and Ras relatives are signal transducers, came the realization that many signaling pathways are more precise, sending controlled and probably weakly amplified signals to specific targets. These intracellular signals are often maintained in macromolecular form rather than being passed to small molecules.

The proteins that carry these signals are not acting in the classic fashion of enzymes that are designed to modify large numbers of substrate molecules. These signal transducers, even if they catalyze an event such as phosphorylation, generally affect small numbers of target molecules and have often separated their catalytic function from their binding regions, which can bring substrates to the catalytic centers, link the signal transducers to upstream proteins, and localize protein complexes to particular cellular subregions. The binding domains are often modular ones constructed with a common core recognition ability coupled to a fine specificity control. They modulate the interaction of proteins with other proteins and therefore determine the paths of signal transduction systems. They are generally controlled, also, so that the aggregates they form are transient, pathways forming only when the signal is being transmitted and then disaggregating when the signal has passed. The signal transduction protein must be highly integrated, with all of the elements working together to send just the appropriate quanta of signal for the specific need.

The preceding paragraphs are generalizations based on scanty and fragmentary evidence. With many transduction pathways known only in outline and more sure to be found, there is a deeply exciting richness yet to be elucidated. But many investigators are now working within the framework presented above, trying to find the relevant units and understand their integration.

Protein–protein interaction has long been studied, but the realization of the importance of defined binding modules came with the recognition that the Rous sarcoma virus oncogene, *src*, has not only a tyrosine kinase catalytic domain but also two other domains that are held in common with proteins that either do or do not have kinase domains. First came the SH2 and then the SH3 domains (for Src homology modules) (Koch et al., 1991), which were later shown to be specific binding domains. Later still, the Pleckstrin homology (PH) domains were recognized (Mayer et al., 1993; Haslam et al., 1993). Many investigators have their computers trained on recognizing other such domains, but no other large, modular families have been reported, suggesting that there may be few left to find.

SH2, SH3, and PH domains have a few common properties. First, they are true protein domains: they form compact units that maintain their structures in isolation, and each has its N- and C-termini in close apposition so that they can be plugged into the surface of proteins. They are not restricted to particular types of signal transduction proteins; they occur in protein kinases, lipid kinases, protein phosphatases, phospholipases, Ras-controlling proteins, and even transcription factors, although they are rarely found in receptors. They are also found in adapter proteins such as Crk, which have no enzymatic function and appear to act solely to aggregate other proteins. SH3 and PH domains are also seen in cytoskeleton proteins, where they may mediate the action of signal transduction pathways on cellular architecture and cell movement. There is no pattern to the number of these domains that occur in proteins, nor in their location within proteins (Figure 1). PH and SH3 domains are found in all eukaryotic organisms, including yeast, while SH2 domains are not found in yeast. None of these domains are seen in prokaryotes (Koch et al., 1991; Mayer and Baltimore, 1993; Pawson and Schlessinger, 1993; Musacchio et al., 1994b).

This review will present our current understanding of how SH2, SH3, and PH domains are constructed, how they function, and how their activities are integrated into the general scheme of signal transduction.

## SH2 Domains

In cells that respond to growth and differentiation factors through receptor tyrosine kinases (RTKs), among the most heavily phosphorylated proteins are the kinases them-

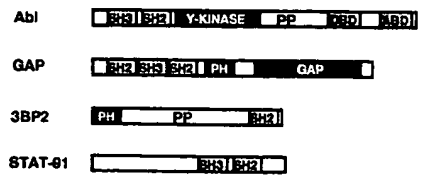



Figure 1. Examples of Proteins with SH2, SH3, and PH Domains

Y-kinase, tyrosine kinase domain; PP, Pro-rich region that binds to SH3 domains; DBD, DNA-binding domain; ABD, actin-binding domain; GAP, GTPase-activating domain.

*Present address: Department of Biology, Rosenstiel Basic Medical Sciences Research Center, Brandeis University, Waltham, Massachusetts 02254.

selves. Autophosphorylation of the kinase is critical for transmitting signals to downstream molecules. In the early 1990s, it became clear that most of the molecules that stably bind to phosphorylated RTKs contain SH2 domains. Subsequent work showed that SH2 domains by themselves stably associate with Tyr-phosphorylated, but not unphosphorylated, RTKs (Mayer and Baltimore, 1993). The question then became, how specific are SH2–phosphoprotein interactions? In other words, can an isolated SH2 domain, immersed in a sea of phosphoproteins, choose a specific binding partner? And if specific, how is the specificity encoded within the phosphoprotein and SH2 domain structures? The answers soon followed.

Different SH2s bind to distinct phosphotyrosine (pTyr)-containing regions of the RTK. Mapping of the binding sites revealed that SH2 specificity is largely determined by the three residues immediately carboxyl to the pTyr, pTyr-X-X-X, where X-X-X varies for different SH2s. For example, a consensus binding site for the two SH2 domains of phosphatidylinositol 3′-kinase (PI3K) is p-Tyr-(Val/Met)-X-Met. Site-directed mutations in the platelet-derived growth factor receptor (PDGFR) at sites containing the consensus sequence—$Y^{740}$MDM and $Y^{751}$VPM—severely disrupt binding of PI3K to PDGFR. These mutations, however, do not affect the affinity of other SH2-containing molecules for the receptor, such as RasGAP (Cantley et al., 1991; Kazlauskas et al., 1992; Pawson and Schlessinger, 1993).

Songyang et al. devised a systematic approach for determining the optimal peptide ligand for specific SH2 domains (Songyang et al., 1993, 1994). They synthesized an 8 amino acid phosphotyrosyl peptide library, randomized at positions Y+1 to Y+3 (the nomenclature refers to the residues carboxyl to the pTyr as +1, +2, etc.). Of the 22 SH2 domains tested, each selected a unique peptide sequence, except for those from Src family proteins, all of which selected the sequence pYEEI (which might suggest that the Src family has redundant SH2 functions). Most SH2 domains fall into one of two broad categories: group I SH2s prefer pTyr-hydrophilic-hydrophilic-hydrophobic; group II SH2s select pTyr-hydrophobic-X-hydrophobic. In proteins that contain two SH2 domains, the SH2s may have either similar or different specificities (Songyang et al., 1993; Panayotou et al., 1993). On the basis of the unique motif recognized by an individual SH2 domain, protein sequence data bases were searched for potential ligands. In some cases, the proteins predicted to interact were found later to do so.

Shoelson and colleagues have quantified the relative binding affinity of various ~12-mer phosphotyrosyl peptides for the N-terminal SH2 domains of the Syp phosphatase and PI3K (Piccione et al., 1993; Case et al., 1994). Using a competition binding assay they demonstrated that the peptides display a continuum of affinities for SH2 domains ranging over three orders of magnitude (pTyr alone binds ~$10^4$-fold less tightly than the highest affinity peptides). When the lengths of the peptides were shortened from 12 amino acids to 6, there was little loss in affinity for the PI3K SH2 domain, as long as the N- and C-termini were blocked (Ac-X-pY-X-X-X-X-$CONH_2$). However, the Syp SH2 domain behaved differently. Similar truncations in Syp-specific peptides severely diminished binding to the SH2 domain. We explore the physical basis for these different behaviors later. Measurements of the dissociation constant ($K_d$) for SH2 domains and high affinity peptide ligands have ranged from <nM to the μM range (Mayer and Baltimore, 1993); it appears that values of 20–100 nM are probably correct.

What is the in vivo significance of different SH2 specificities? Work by Pawson and coworkers begins to answer this question (Marengere et al., 1994). By changing a single residue in the Src SH2 domain, they switched its selectivity toward that of a SEM-5 SH2 domain. When the altered SH2 domain was spliced into the *sem-5* gene, it provided SEM-5-like activity in Caenorhabditis elegans better than did a construct containing the Src SH2, indicating a correlation of binding specificity with biological activity.

Therefore, although all SH2 domains bind to phosphotyrosyl proteins, they are not promiscuous in choosing partners. Ligand specificity is largely determined by the linear sequence surrounding the pTyr. However, these specificities are not absolute, and there may be more than one SH2 within a cell with high affinity for a particular ligand. Therefore, in vivo, the ability of an SH2 domain to engage a particular phosphoprotein may depend critically on the local concentration of proteins as well as the modulating effect of other domains found on interacting proteins. We take up these complications in later sections. In the section that follows, we discuss the structural basis for SH2 specificity.

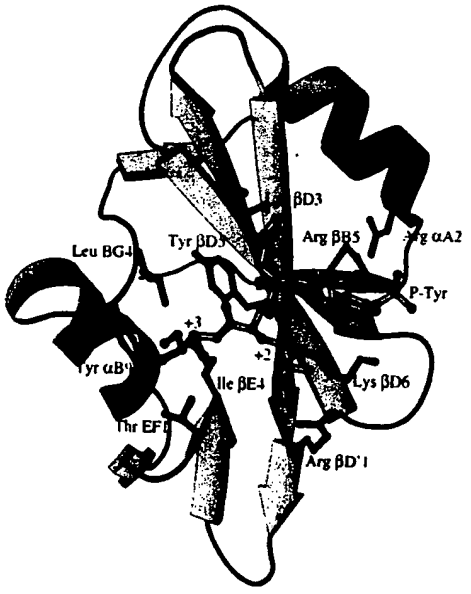



Figure 2. Schematic Diagram of the Src SH2 Domain, Illustrating the Orientation of the Bound Phosphopeptide

The phosphopeptide backbone has been highlighted in yellow, while some SH2 residues in the ligand-binding site have been highlighted in orange and blue (see text for details). This figure is based on a figure that appeared in Waksman et al. (1993) and was redrawn by Chi-Hon Lee and John Kuriyan.

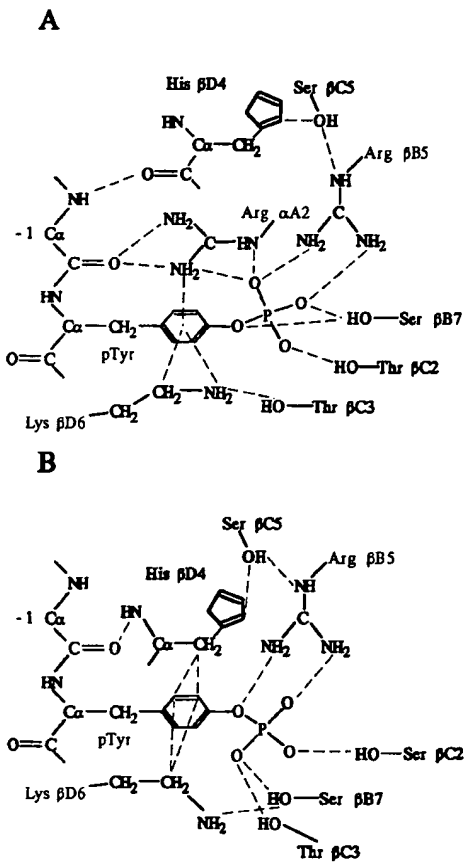



Figure 3. pTyr-Binding Sites

The Src pTyr-binding site (A) and the Syp pTyr-binding site (B). SH2 domain and phosphopeptide residues have been drawn in black and blue, respectively. Residues that are mentioned in the text have their names rendered in red, while noncovalent interactions are indicated by broken red lines (see text for details). For the sake of clarity, not all contacts between the pTyr and SH2 domain have been included. These figures are based on figures that appeared in Waksman et al. (1993) and Lee et al. (1994) and are reproduced by permission of the authors and publishers.

## SH2 Structure

All SH2 domains solved to date contain a large central antiparallel β sheet, two flanking α helices, and follow a general pattern of β-α-β-β-β-β-α-β. Given the similarities in SH2 secondary structure, a nomenclature was adopted to facilitate comparisons (Eck et al., 1993): the β sheets are lettered βA, βB, et cetera; the α helices are αA and αB; and loops are labeled by the 2° structures they connect. Residues are numbered consecutively within the 2° structures. Figure 2 shows a schematic diagram of the Src SH2 domain bound to a high affinity phosphopeptide and illustrates the nomenclature.

### *The pTyr-Binding Site*

SH2 domains bind specifically to pTyr-containing proteins and show little or no affinity for phosphoserine/threonine sequences. Structures of SH2 domains bound to pTyr peptides give a structural basis for this preference (Waksman et al., 1992). The pTyr side chain projects into a deep pocket. At the bottom of the crevice, the negatively charged phosphate is coordinated by a bidentate interaction with the positively charged side chain of Arg βB5 (Figure 3). Since Arg βB5 is 7.5 Å removed from the α-carbon atom of the pTyr (in the Syp structure) (Lee et al., 1994), it is inaccessible to the shorter side chains of phosphoserine or Thr. This Arg is invariant, and mutation of the Arg, even to Lys, abolishes the ability of the Abl SH2 domain to recognize phosphoproteins (Mayer and Baltimore, 1993).

The pTyr recognition site also contains an extensive network of hydrogen bonds to the phosphate and a number of hydrophobic contacts to the phenol ring. While one might think the architecture of the pTyr-binding site is strictly conserved, there is in fact some flexibility in its design. For Src and Lck, there are two amino-aromatic contacts to the phenol ring (one is supplied by Lys βD6 and the other by Arg αA2; Figure 3A) (Eck et al., 1993; Waksman et al., 1993). Bonds between positively charged atoms and benzene rings have recently been described in a number of protein structures (Waksman et al., 1993; Miller, 1994); therefore, it was natural to think that these two amino-aromatic interactions would be critical for pTyr recognition. Apparently, at least for some SH2 domains, this is not the case. Syp is rare among SH2 domains in lacking Arg αA2. Instead, it has a Gly. Furthermore, in the Syp structure, Lys βD6, which supplied the other amino-aromatic contact in Src, contacts the phenol ring solely with its hydrocarbon chain and not with the amine (Figure 3B). Despite the loss of the amino-aromatic interactions there are no major structural alterations in Syp's pTyr pocket (although Arg βB5, which binds to the two terminal oxygens in Src, binds to the phenolic oxygen and a terminal phosphate oxygen in Syp; Figure 3) (Lee et al., 1994). A further demonstration of the plasticity of the pTyr-binding site is seen in the nuclear magnetic resonance (NMR) structure of the SH2 domain of PLC-γ1, where the BC loop, which makes important hydrogen-bond contacts in the Src, Lck, and Syp structures, no longer directly contacts the phosphate (PLC-γ may replace these contacts with stronger associations between the phosphate and positively charged residues found in the PLC-γ pocket) (Pascal et al., 1994). Thus, although certain residues may contribute greatly to pTyr-binding interactions in a given SH2 domain, they are replaceable in others. Within the SH2 domain kingdom, only Arg βB5 is invariant.

### *The Peptide-Binding Site: Act I (Enter Src and Lck)*

The first structure of a ligated SH2 domain was solved using a low affinity peptide (Waksman et al., 1992). In this structure, there were few contacts to the peptide chain. Now that there are structures of both low and high affinity ligands complexed to SH2 domains, it is clear that what differs between these ligands is not how they bind to pTyr, which is relatively unaltered, but rather the extent of interactions to the peptide chain. High affinity ligated structures were determined for Src and Lck bound to an 11-mer peptide, derived from hamster middle T antigen, that contained the Src family–binding motif pYEEI (Eck et al., 1993; Waksman et al., 1993). Four residues in the peptide, pYEEI, make much more extensive contact with the protein than the other seven. Of these four residues, the most

elaborate pockets cradle the pTyr and Ile +3. Binding of this peptide to the Src SH2 domain has therefore been described as "a two-pronged plug engaging a two-holed socket" (Waksman et al., 1993). The Ile +3–binding pocket is lined, as expected, primarily with hydrophobic residues (see Figure 2). The EF and BG loops form part of the Ile +3–binding pocket and have been described as a "set of jaws" that clamp down on the Ile +3 residue (Eck et al., 1993).

The glutamate residues (Y+1 and Y+2) are both solvent-exposed in the Src and Lck structures, and there is no obvious glutamate-binding pocket. There are positively charged residues in the vicinity of the glutamates that may contribute to selectivity (e.g., Lys βD3, Lys βD6, and Arg βD'1; see Figure 2). However, some of these residues are also found on SH2 domains that select for hydrophobic amino acids at the +1 and +2 sites. Moreover, the distance from these residues to the glutamates is too large for there to be a strong interaction. This is somewhat disturbing, since for Lck, the selectivity for a Glu at the +1 site is as high as it is for Ile +3 (Songyang et al., 1993). Furthermore, in the PDGF receptor, a likely binding site for Src is pY$^{579}$I-pY$^{581}$V (where both Tyrs are phosphorylated), while for the CSF-1 receptor, the site may be Y$^{559}$TFI (Mori et al., 1993). Both of these sites significantly deviate from the putative consensus, pYEEI. Careful biochemical analysis and systematic mutations of the consensus ligand motif and SH2 domain are needed to assess better the contribution of individual residues toward binding specificity.

### *The Peptide-Binding Site: Act II (Enter Syp and PLC-γ1)*

Recently, the Syp and PLC-γ1 SH2 structures were determined (Lee et al., 1994; Pascal et al., 1994). In contrast with Src and Lck, which are classified as group I ligand binders (pTyr-hydrophilic-hydrophilic-hydrophobic), Syp and PLC-γ1 are group II binders (pTyr-hydrophobic-X-hydrophobic). What is most strikingly similar about these four structures, besides the obvious pTyr-binding site, is the manner in which the peptide is held by the protein. A direct comparison of the Syp and Src ligated peptides reveals that when the peptides' α-carbon atoms are compared, the backbone conformation from the pTyr to the +3 position are virtually identical (Lee et al., 1994). This conservation had been predicted by Harrison and coworkers, who had noted in the Lck structure that the extended peptide–protein interface contained extensive contacts to the peptide's backbone that "suggest tight binding will generally require the peptide conformation seen in this structure" (Eck et al., 1993).

What was unexpected in the new structures is the length of the peptide that contacts the protein. For Syp, there are contacts out to the +5 position, and for PLC-γ1 they extend out to +6. Furthermore, the binding site for the peptide in both of these structures is a shallow groove—in contrast with the "two-holed socket" of Src and Lck. A number of features account for the different binding pockets. All group I SH2s (Src and Lck) contain bulky side chains at position βD5 (Tyr or Phe) (Songyang et al., 1994). In contrast, group II SH2s (Syp and PLC-γ1) contain smaller, nonaromatic side chains at βD5. Tyr βD5 acts as the divider between the pTyr and Ile +3–binding pockets (see Figure 2). In its absence, a slender channel is created between the two pockets. This groove is hydrophobic and provides to hydrophobic residues at the +1 position a place to bury their side chains partially.

Further selectivity for hydrophobic residues by the Syp and PLC-γ1 SH2 domains is imparted by the BG loop. The conformation and length of this loop varies among the SH2 domains, as does the extent to which it contacts the ligated peptides. Therefore, this loop may play an important role in determining peptide selectivity. In the Src and Lck structures, the BG and EF loops seal off the Ile +3-binding pocket. This action forces the peptide chain away from the protein and precludes significant interactions beyond Ile +3 (see Figure 2). In the Syp and PLC-γ1 structures, these loops are less constraining, and a groove opens up beyond the +3 site, which allows for the more extended contacts to the peptide chain.

Consistent with the structural studies, mutation of the +5 residue in the Syp peptide results in a >100-fold decrease in the peptide's affinity. This result explains why truncation of Syp-specific peptides to 6-mers (mentioned above) severely diminishes binding to the protein. Similar biochemical studies for PLC-γ1 suggest that the extended contacts (+5 and +6) are not critical for high affinity interactions (Case et al., 1994). It is unclear why these results and the NMR structure are at odds. Others have reported that residues on the N-terminus of the pTyr alter the binding affinity of peptides to some SH2 domains (Bibbins et al., 1993; Nishimura et al., 1993). These results suggest that although much has been learnt about SH2 domain specificity, there is still more to go. The future will no doubt see more structures, more bound ligands, and, hopefully, quantitative analysis of the binding interactions between whole phosphoproteins and SH2 domains.

### *Kinetics of Ligand Binding*

High affinity phosphopeptides bind to SH2 domains rapidly: $k_{on}$, ~1 × $10^5$ $M^{-1}s^{-1}$ to 2 × $10^6$ $M^{-1}s^{-1}$, but dissociation is also moderately fast, even for high affinity ligands: $k_{off}$, ~0.1 $s^{-1}$ ($t_{½\ dissociation}$ = 5–10 s) (Panayotou et al., 1993). Since the response time of cells to growth factors is minutes to hours, even after SH2-containing proteins have engaged an RTK, the signal emanating from the receptor may remain open to the modulatory effect of phosphatases as well as competing SH2-containing proteins. Interestingly, the difference between high and low affinity ligands appears to be due to differences in $k_{on}$, with little variation in $k_{off}$ (Felder et al., 1993; Panayotou et al., 1993; Marengere et al., 1994). This seems surprising in light of the crystal structures, which show that high affinity ligands make more extensive contacts to the protein than do low affinity ligands. However, binding to the SH2 domain by a phosphopeptide may be a two-step process. In the first step, the phosphopeptide may become weakly associated with the protein to form an unstable intermediate complex. The rate-limiting transition state may occur subsequently, as the pTyr becomes tightly locked into its binding pocket. Therefore, during dissociation of the peptide, the rate-limiting step could be disengagement of the pTyr. Since the pTyr is held in the binding pocket in a similar manner

for both high and low affinity ligands, $k_{off}$ would be similar for both ligands.

***Do SH2 Domains Bind In a Non-pTyr-Dependent Manner?***

A number of reports claim that SH2 domains interact with proteins in a non-pTyr-dependent manner (Muller et al., 1992; Bibbins et al., 1993; Cooper and Kashishian, 1993). In at least one case, the interaction described (binding of the SH2 domains from the noncatalytic p85α subunit of PI3K to the catalytic p110 subunit) could not be reproduced by another lab and subsequently has been ascribed to a different region of p85α (Hu and Schlessinger, 1994; Klippel et al., 1994). Furthermore, the pTyr-independent interactions described so far are weaker than SH2–pTyr-dependent interactions (Muller et al., 1992; Bibbins et al., 1993). However, if the pTyr-independent interactions occur intramolecularly, then even relatively weak interactions could be used by a protein as a means of allosteric regulation. Therefore, the possibility that some SH2 domains have an effector region outside the pTyr-binding site is an intriguing but at this time speculative idea.

## SH2 Domain Function

How does the binding of an SH2-containing protein to a Tyr-phosphorylated protein help transmit a signal? First, binding may alter the subcellular localization of a protein, bringing it closer to its substrate, or closer to a protein that modifies it. Second, binding may induce conformational changes that alter the catalytic activity of the interacting proteins. SH2-containing proteins use both of these methods.

***Propagation of Signals by a Change In Subcellular Localization***

The RTK-mediated Ras activation pathway has been the paradigm for activation by localization (Pawson and Schlessinger, 1993). Ras, a GTPase associated with the plasma membrane, is activated by the guanine nucleotide exchange factor SOS. Even in quiescent cells, SOS is found in the cytoplasm engaged with the adaptor molecule Grb2 via the latter molecule's two SH3 domains. Activation and autophosphorylation of the epidermal growth factor receptor (EGFR) creates a binding site for the SH2 domain of Grb2 and recruits the Grb2–SOS complex to the membrane, bringing SOS into proximity with its Ras substrate. At the membrane, SOS may be Thr- or Ser-phosphorylated. However, the intrinsic catalytic activity of SOS in activated or quiescent cells is indistinguishable, leaving the biological relevance of SOS phosphorylation unknown (in yeast, phosphorylation of a Ras exchanger, Cdc25, by a cAMP-dependent kinase releases Cdc25 from the membrane; Rozakis-Adcock et al., 1993). Therefore, in EGF-treated cells, Ras activation is due to the increased local concentration of SOS. Localization is also important for the pathways that activate PLC-γ1, PI3K, and RasGAP, because like SOS, these molecules are cytosolic, and their substrates, phospholipids and Ras, are membrane bound.

A variation on the EGFR/Grb2 pathway is used by PDGFR. PDGFR does not bind directly to Grb2 but instead contains a binding site, $pY^{1009}$, for the tyrosine phosphatase Syp (Lechleider et al., 1993; Li et al., 1994). Syp, upon binding to the PDGF receptor, is Tyr-phosphorylated, and it is thought that phosphorylation of Syp provides a docking site for the Grb2–SOS complex. In PDGFR mutants that carry multiple Tyr to Phe mutations, the presence of $Y^{1009}$ is sufficient but not necessary for Ras activation by PDGF (Valius and Kazlauskas, 1993). Therefore, Syp-independent pathways for Ras activation must also exist. There is also genetic evidence that the PDGFR/Syp/Grb2/SOS pathway is physiologically relevant. In *Drosophila melanogaster*, the Syp homolog, corkscrew, positively transmits signals from the putative RTK, torso, on a pathway that mediates cell fate determination for the head and tail regions of Drosophila embryos. SOS and Ras are also essential for this pathway. The ability of Syp to act as an adaptor molecule and bring Grb2 to the kinase may explain the apparent paradox of a phosphatase aiding the signaling pathway of a kinase (Li et al., 1994). The function of Syp's phosphatase activity on this pathway, however, remains a mystery.

Grb2 appears to be (almost) a universal adaptor protein in tyrosine kinase signaling pathways. Besides Syp and EGFR, it also binds to Tyr-phosphorylated forms of Shc and insulin receptor substrate 1 (IRS-1) (Pawson and Schlessinger, 1993). However, phosphorylated Shc and IRS-1 do not stably associate with the membrane, which suggests that they may use a presently unknown mechanism to bring about Ras activation.

***Phosphorylation of SH2-Containing Proteins***

The binding of an SH2-containing protein to a tyrosine kinase increases the likelihood that the SH2-containing protein is going to get phosphorylated (kinases that modify proteins frequently stably associate with their substrates, e.g., rhodopsin kinase and the stress-activated protein kinase family; Lefkowitz, 1993; Dérijard et al., 1994; Kyriakis et al., 1994). Phosphorylation of SH2-containing molecules may recruit other molecules to the activated complex; e.g., phosphorylation of Syp by PDGFR created a new binding site for Grb2. Alternatively, phosphorylation may alter the intrinsic catalytic activity of the SH2-containing molecule; e.g., EGFR phosphorylation of PLC-γ1 increases the catalytic activity of PLC-γ1 toward profilin-bound phosphatidylinositol (4,5)-bisphosphate ($PtdInsP_2$; hydrolysis of $PtdInsP_2$ frees profilin, which then regulates actin polymerization) (Aderem, 1992).

The JAK/STAT pathway provides another rich example of how phosphorylation of an SH2-containing molecule is used in signal transduction pathways (Darnell et al., 1994). JAKs are members of a family of nonreceptor tyrosine kinases, the Janus kinases, that contain two putative tyrosine kinase domains. Their name is derived from that of the Roman god, Janus, god of gates and doorways and the porter of heaven; he is often portrayed with two faces looking in opposite directions (thus, the first month of the year is January). JAKs associate with membrane-bound receptors, e.g., PDGFR and the interferon α receptor. When a cell is treated with growth factors or cytokines, JAK family members may be activated and then phosphorylate members of the STAT protein family. Phosphorylated STATs form homo- and probably heterodimers, held to-

gether by a reciprocal interaction of the SH2 domain on one STAT with a pTyr on the opposing STAT. Unphosphorylated STATs are cytoplasmic, but upon phosphorylation and dimerization they, along with other molecules that bind to the dimer, translocate to the nucleus. In the nucleus, the complex binds to specific DNA sites and activates transcription. STATs are an example of transcription factors regulated by Tyr phosphorylation.

***IRS-1: Phosphorylation of a Signaling Molecule That Does Not Contain an SH2 Domain***

In insulin-stimulated cells, a major substrate for the insulin receptor is the cytosolic protein IRS-1. It is not entirely clear how IRS-1, which lacks an SH2 domain to direct it to the membrane, gets phosphorylated. The insulin receptor kinase is specific for certain IRS-1 Tyr-containing motifs (Shoelson et al., 1992). However, in general, tyrosine kinases are not as specific as serine and threonine kinases. Moreover, IRS-1 has 21 potential Tyr phosphorylation sites, at least eight of which can be phosphorylated (Myers et al., 1994). Whether the insulin receptor is specific for all of these phosphorylation sites is not known. In any event, once IRS-1 is phosphorylated it associates with a number of proteins, including PI3K, Grb2, Syp, and Nck. Because IRS-1 interacts with so many proteins, it has been dubbed a "docking protein". The docking protein links the insulin receptor to many SH2-containing signaling pathways, and in this respect it is analogous to the cytoplasmic tail of the PDGFR. In these signaling complexes, many opportunities might exist for interactions between neighboring molecules. However, to date, there is little evidence for such interactions, and it is not even known whether signaling complexes containing proteins from multiple pathways generally form under physiological conditions.

***Allosteric Activation of SH2-Containing Proteins***

Both the Syp phosphatase and PI3K show enhanced catalytic activity upon binding to phosphopeptides (Carpenter et al., 1993; Shoelson et al., 1993; Sugimoto et al., 1994). For Syp, the increase in activity can be more than 50-fold; for PI3K it is about 5-fold. While it is conceivable that phosphoprotein binding induces a conformational change in the SH2 domain that is then transmitted to the catalytic domain, the crystal structures suggest that this is unlikely. Very little structural alteration is seen within SH2 domains upon ligand binding (Lee et al., 1994). However, although SH2s may not be materially altered by binding, their relation to the rest of the protein could be changed. Thus, an internal interaction of the SH2 with the protein could be modified when the SH2 binds to an exogenous ligand. The best example of this comes from the Src family, where a pTyr residue in the carboxyl tail, $pY^{527}$, interacts with the Src SH2 domain to regulate the kinase negatively (Superti-Furga et al., 1993). A somewhat similar SH2/carboxyl tail repression mechanism might be used by the Syp phosphatase as limited protease digestion of Syp's carboxyl tail activates the phosphatase and mitigates the increase in activity seen upon addition of phosphopeptides (Sugimoto et al., 1994). Both Syp and PI3K contain two SH2 domains, and there is evidence that engagement of both domains is required for maximal stimulation of the catalytic domain.

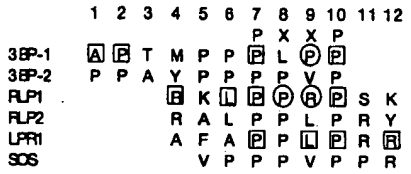



Figure 4. SH3 Peptide Ligands

3BP-1, 3BP-2 , RLP1, and RLP2 are class I ligands, while LPR1 and the SOS-derived peptide are class II ligands (see text for details). Regions of these ligands that are contained in the PPII helix in ligated structures have been highlighted in yellow (a structure for the ligated form of the SOS-derived peptide has not been determined, so this peptide is not shaded). Residues that have been identified as critical for high affinity interactions with SH3 domains have been boxed, while residues that contribute less to binding have been circled (thorough mutagenic analysis has not yet been done for 3BP-2, RLP2, and the SOS-derived peptide, so these peptides have not been marked).

**SH3 Domains**

While the biological functions of SH2 domains are fairly well understood, understanding the biology of SH3 domains has lagged behind. A key for deciphering the function of SH3 domains came when two SH3-binding proteins, 3BP1 and 3BP2, were identified by screening a cDNA expression library with the Abl SH3 domain (Cicchetti et al., 1992). The location of the possible binding regions on 3BP1 and 3BP2 was narrowed to ~10 amino acid Pro-rich sequences (Figure 4) (Ren et al., 1993). Since then, a number of in vivo SH3 ligands have been identified, e.g., SOS, PI3K, and $p47^{phox}$, and the binding sites have similarly mapped to Pro-rich sequences (Finan et al., 1994; Musacchio et al., 1994b). Other high affinity SH3 peptide ligands were identified by screening with a chemically synthesized peptide library (Chen et al., 1993; Yu et al., 1994). Schreiber and coworkers started with a biased random peptide library, XXXPPXPXX, where X is any amino acid other than Cys, and found that both the Src and PI3K SH3s selected for two classes of ligands (the bias was based on residues found in 3BP1 and 3BP2; without the bias, no high affinity ligands were found). Class I ligands are RXLPPδPXX, where δ is L for Src and R for PI3K. Class II are XXXPPLPXR for both Src and PI3K. In class I ligands for Src and PI3K, an Arg is N-terminal to the Pro-rich sequence, while for class II ligands it is C-terminal. Examples of class I and II ligands used in structural studies are listed in Figure 4. Mutational analysis has identified residues that are particularly important for binding, and these residues are boxed in Figure 4; less important residues have been circled (Ren et al., 1993; Yu et al., 1994; Feng et al., 1994). All high affinity SH3 ligands identified so far contain a PXXP motif, and the sequences in Figure 4 have been aligned by this motif. In the following section, we consider the structural basis for SH3 selectivity.

**SH3 Structure**

Structures for eight SH3 domains have been reported, four of these with ligands (PI3K, Abl, Fyn, Src) (Feng et al., 1994; Musacchio et al., 1994a; Yu et al., 1994). Although there are variations among these structures, the overall topology is well conserved. The basic fold consists

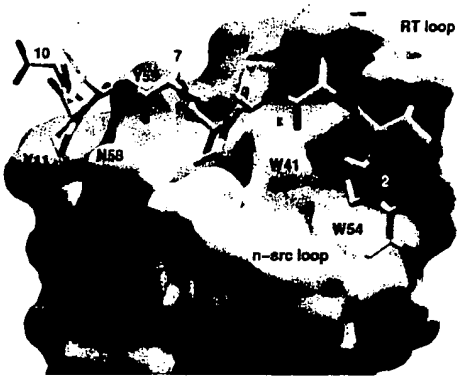



Figure 5. The 3BP-1 Peptide Bound to the Abl SH3 Domain

The numbering on the residues of the peptide corresponds to the numbering for the 3BP-1 peptide in Figure 4. Residues 4–10 of the peptide adopt a PPII-helical conformation (see text for details). Pepide color coding: nitrogen, blue; oxygen, red; and carbon, white. Some residues and regions of the Abl SH3 domain that define the ligand-binding surface are also highlighted (see Musacchio et al., 1994a, for a more complete description of the interactions). This figure previously appeared in the paper by Musacchio et al. (1994a) and is reproduced by permission of the authors and publisher.

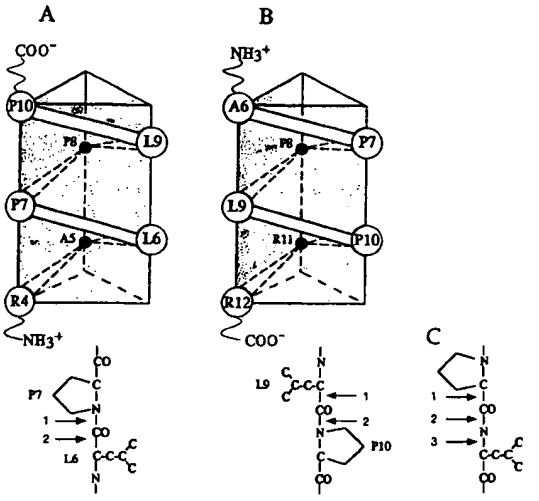



Figure 6. Schematic Diagram Illustrating PPII Helices for Src-Binding Peptides

RLP2 (A) and LPR1 (B). Below each illustration is a stick figure for two residues that are found in each helix. Notice that two backbone bonds (numbered in red) separate the Pro and non-Pro side chains, while in (C), three backbone bonds separate the side chains (see text for details).

of five antiparallel β strands (named "A" to "E") that pack to form two perpendicular β sheets. A hydrophobic patch that contains a cluster of conserved aromatic residues and is surrounded by two charged and variable loops (the AB and BC loops) forms the ligand-binding pocket.

The first ligated structure was of the PI3K SH3 bound to the peptide RLP1 (see Figure 4; Yu et al., 1994). As with the SH2 domain, very little structural alteration in the SH3 is detected upon ligand binding. The central portion of this peptide, residues Arg4 to Pro10, adopts a polyproline type II (PPII) helix (the numbering of these residues refers to the numbers shown above the peptide sequences in Figure 4). PPII helices appear in all bound SH3 ligands examined so far. Figure 5 illustrates the PPII helical conformation of the 3BP-1 peptide bound to the Abl SH3 domain (the PPII helical regions of other SH3 ligands have been highlighted in yellow in Figure 4). Though the PPII motif was first identified in 1955 in crystals of L-polyproline, their ubiquity in globular proteins has only recently been recognized (Adzhubei and Sternberg, 1993). In contrast with their more renowned counterparts, the $3_{10}$ and α helices, which are predominantly right handed, PPII helices are left handed. The number of residues per turn in a perfect PPII helix is ~3.0; therefore, we can imagine the PPII portion of the bound peptide as a trigonal prism (Figure 6A is drawn for the Src-binding peptide RLP2). Two edges of the prism contact the protein (the SH3 domain would be where your face is) (Yu et al., 1994). This agrees well with the critical residues identified on RLP1 by mutational analysis. In globular proteins, most of the amide nitrogens and oxygens in the PPII helix form hydrogen bonds to water. PPII helices therefore tend to be solvent-exposed and mobile, ideal sites for interactions with other proteins. No hydrogen bonds between the peptide backbone and SH3 domain were seen in the PI3K structure, and only two hydrogen bonds were seen in the Abl and Fyn structures (Musacchio et al., 1994a). Therefore, the driving force for bonding comes primarily from interactions between the ligand side chains and the protein.

Although Pro residues are not strictly required for PPII helices (~25% of PPII helices in globular proteins do not contain Pro [Adzhubei and Sternberg, 1993]), the propensity of Pro to adopt this conformation readily explains their presence in SH3 ligands (SH3-binding, Pro-rich peptide ligands may exist as PPII helices free in solution; Musacchio et al., 1994b). The two critical Pros in the PXXP motif directly pack against the SH3-binding protein, lodging themselves into a host of conserved, mostly aromatic, residues. The N-terminal Arg found on the class I ligands RLP1 and RLP2 (Arg4; see Figure 4) engages in a critical electrostatic interaction with Asp-99 (numbering is for Src); mutation of either Arg4 or Asp-99 severely disrupts ligand binding (Yu et al., 1994). However, Asp-99 also binds the C-terminal Arg (Arg-12) in class II ligands LPR1 and SOS. Therefore, the orientations of class I and II ligands on Src are opposite! Compare Figures 6A and 6B (Feng et al., 1994). Furthermore, the positions of the critical Pros on the face of the trigonal prism are exchanged with the non-Pro residues. This exchange keeps the distance between the prolyl ring and the non-Pro side chain constant (notice that in the stick figures in Figures 6A and 6B the two side chains are separated by two backbone bonds; if you change the position of the Pro and non-Pro within a stick figure, the side chains become separated by three backbone bonds; e.g., compare Figures 6B and 6C). Apparently, the SH3-binding pocket is flexible enough to accommodate both ligand orientations, and the tightest-binding class I and class II ligands have similar affinities for the SH3 of PI3K (8 μM and 13 μM, respectively; Chen et al., 1993). The

lack of substantial hydrogen bonding to the peptide backbone may further contribute to the orientation androgyny. Whether either orientation predominates in nature remains to be determined; the SOS peptide (see Figure 4) is likely to interact with Grb2 in a class II interaction (Feng et al., 1994), while the 3BP1–Abl interaction is class I oriented (Musacchio et al., 1994a).

Like that of SH2, the specificity of SH3 is thought to be determined by the variable amino acids that surround invariant residues. This specificity is evident in that Abl SH3 does not bind well to Src-specific ligands, while Src, in turn, binds poorly to Abl-specific peptides (Rickles et al., 1994). However, Asp-99, the residue that may dictate Src's specificity for Arg4, is found in virtually every SH3 domain (except Abl, which has a Thr). Moreover, the Fyn SH3 domain, which selected for a Src-like peptide (RXLPXXP) in a phage display library (Rickles et al., 1994), has a residue homologous to Asp-99, yet binds to the Abl-specific 3BP-2 peptide with fairly high affinity, ~35 mM (Musacchio et al., 1994a). Therefore, specificity could be influenced by residues lying outside the PPII helix; for instance, in the Abl–3BP1 and Fyn–3BP2 structures, Pro2 (which lies outside the PPII helix; see Figure 4) makes significant contact with the SH3 domain (Musacchio et al., 1994a). However, if longer peptides are needed, then perhaps in vivo specificity will depend upon nonlinear sequences, i.e., tertiary structure. There is evidence that disparate regions of the human inmmunodeficiency virus (HIV) protein Nef mediate interactions with the SH3 domain of Hck (Saksela et al., 1995). Nonlinear sequences might mimic the PPII helix and also provide additional regions that interact with the SH3 domain. Alternatively, SH3 domains may carry only crude specificity constraints, while the fine tuning of protein–protein interactions could be determined by other modular domains on the interacting proteins.

### Interactions between SH2 and SH3 Domains

SH2 and SH3 domains are found together on many proteins; might their activities be coordinated? There is evidence to suggest so. For instance, the c-Src kinase is negatively regulated by interactions between its SH2 domain and its Tyr-phosphorylated C-terminal tail. However, this inhibition also requires a functional SH3 domain. Mutation of the SH3 activates the kinase and increases the accessibility of the SH2 domain to exogenous substrates (Superti-Furga et al., 1993). In addition, a phosphoprotein that associates with Src during mitosis, p68, has been isolated. p68 interacts with both the isolated Src SH2 and the SH3 domains, but neither domain competes for p68 as well as the tandem SH2–SH3 module (Fumagalli et al., 1994; Taylor and Shalloway, 1994). A similar synergism is observed in the binding of the 110 kDa actin filament–associated protein to Src's SH2 and SH3 domains (Flynn et al., 1993).

Structural evidence in support of SH2–SH3 cooperativity comes from Eck et al. (1994), who crystallized a fragment containing the Lck SH2–SH3 domains (Lck is a Src family tyrosine kinase member). In the Lck structure there are few intramolecular contacts between the SH2 and SH3 domains; however, the fragment crystallized as a dimer and there are extensive intermolecular SH2–SH3 contacts. Interestingly, an SH2 Pro lies in the binding pocket of the opposing SH3 domain and is bound in an orientation similar to that of a PXXP Pro. The Lck SH2–SH3 fragment was also crystallized in the presence of a phosphorylated peptide that corresponds to Lck's carboxyl tail. While the peptide's pTyr binds in an orientation similar to that observed in SH2–peptide complexes, the peptide chain does not; it lies in the crease of the intermolecular SH2–SH3 interface. This result is intriguing, and if c-Src exists as a dimer in vivo, it may explain why both a functional SH2 and an SH3 are required for the negative regulatory effect of Src's phosphocarboxyl tail. To date, there is no evidence that c-Src does exist as a dimer in solution but, as Eck et al. point out, in vivo, Src is a membrane-bound protein, and membrane localization may favor dimer formation. Moreover, membrane-bound Src may further be concentrated within the membrane at signaling center complexes (Lisanti et al., 1994). The consequences of cellular organization on signal transduction pathways has been a largely unexplored area of cell biology.

### SH3 Function

Mutational analysis first demonstrated the biological importance of the SH3 domain: mutation of the SH3 domain of the adaptor molecule v-Crk mitigates this protein's transforming potential. For the nonreceptor tyrosine kinases Abl and Src, deletion of the SH3 domain activates the transforming potential of the proto-oncogene products. The importance of SH3 domains in signal transduction pathways was highlighted in C. elegans, where mutation of either of SEM-5's two SH3 domains blocked vulval development (Mayer and Baltimore, 1993). Although our understanding of the physiological role of SH3-mediated interactions is still at an early stage, recent studies suggest various places where SH3 interactions mediate critical protein–protein interactions. These interactions are used to organize protein complexes within the cell, bring substrates to enzymes, and regulate enzymatic activities, as discussed in the following sections.

#### *Role of SH3 in Cellular Localization*

Compartmentalization of proteins plays an important role in the regulation of signal transduction processes. Many of the proteins that localize to the plasma membrane or the cytoskeleton contain SH3 domains (e.g., the actin-binding protein α-spectrin, nonmuscle myosin Ib, and the S. cerevisiae protein ABP-1), suggesting that SH3 domains mediate localization to these regions (Mayer and Baltimore, 1993). Using a microinjection approach, Bar-Sagi et al. demonstrated that the SH3 domain of PLC-γ is responsible for the targeting of that protein to cytoskeletal microfilaments, while both SH3 domains of Grb2, but not its SH2 domain, are required for Grb2 localization to membrane ruffles (Bar-Sagi et al., 1993). Although many SH3-containing molecules associate with actin filaments, SH3 domains probably do not directly bind to actin. It is also likely that SH3-mediated interactions between α-spectrin and the Pro-rich C-terminal tail of the amiloride-sensitive $Na^+$ channel dictates that channel's localization to the apical

membrane of polarized epithelial cells. Since many ion channels contain Pro-rich sequences, this mechanism of localization may be generally applicable to ion channels (Rotin et al., 1994).

There are also many examples of SH3-mediated interactions influencing signaling pathways containing G proteins of 21 kDa. For example, the Grb2–SOS interaction modulates Ras activity (see above), and the C-terminal SH3 domain of Grb2 binds the N-terminal SH3 domain of Vav, a hematopoietic-specific guanine nucleotide exchange factor (Ye and Baltimore, 1995). In S. cerevisiae, genetic analysis suggests that the SH3-containing protein Bem1p, which regulates cell polarity, interacts with the guanine nucleotide exchange factor BUD5 and a GTPase-activating protein, Bem2p. In addition, some guanine nucleotide exchange factors (e.g., Vav and CDC25) and guanosine triphosphatase–activating proteins (e.g., RasGAP) contain their own SH3 domains (Mayer and Baltimore, 1993). While the frequency of SH3 domains on small GTP-binding protein pathways suggests a common mechanism of regulation, what that mechanism might be is not known.

SH3-mediated interactions also play a role in the assembly of the phagocyte NADPH oxidase system. Phagocyte NADPH oxidase catalyzes reduction of molecular oxygen to superoxide. The NADPH oxidase consists of a membrane-bound flavocytochrome b, composed of two subunits, gp91$^{phox}$ and p22$^{phox}$. Activation of the oxidase occurs when a small G protein, p21$^{rac}$, and three cytosolic SH3-containing proteins, p47$^{phox}$, p67$^{phox}$, and p40$^{phox}$, translocate to the membrane. The two SH3 domains of p47$^{phox}$, both of which are required for the assembly of the oxidase complex, bind to p22$^{phox}$ and p67$^{phox}$ (Sumimoto et al., 1994). In quiescent cells, the p47$^{phox}$ SH3 domains may be blocked by an intramolecular interaction. They are unmasked in vitro by activators of the oxidase, such as arachidonic acid and sodium dodecyl sulfate. Interestingly, the point mutation P156Q, found in a patient with chronic granulomatous disease, occurs in a Pro-rich region of p22$^{phox}$ and abolishes interaction with the p47$^{phox}$ SH3 domain. The C-terminal SH3 domain of p67$^{phox}$ also binds to p47$^{phox}$ (Finan et al., 1994). These results reveal that a complex SH3 network directs the assembly of the phagocyte NADPH oxidase system. Apparently, in this system, SH3-mediated interactions are regulated in contrast with the Grb2/SOS SH3–mediated contacts, which are constitutive.

***Role of SH3 in Regulating Enzymatic Activities***

As was previously mentioned for c-Src, SH3-mediated interactions negatively regulate the kinase activity of nonreceptor tyrosine kinases. The Abl tyrosine kinase is also negatively regulated by its SH3 domain (Mayer and Baltimore, 1993). Although c-Src is activated by mutations in the SH3 domain, in the context of a constitutively active variant of c-Src, Y527F, mutations in the SH3 domain *decrease* transforming activity. However, in another activated variant of c-Src, E378G (a mutation within the kinase domain), the SH3 domain is not required for the transforming activity (Seidel-Dugan et al., 1992). It is not clear why there are different requirements for the SH3 domains in different contexts.

An SH3-mediated interaction is also implicated in the activation of dynamin (Gout et al., 1993). Dynamin is a 100 kDa microtubule-activated GTPase that plays a critical role in the initial stages of endocytosis. Dynamin's GTPase activity is activated by the binding of SH3 domains from Grb2, Src, Fgr, and Fyn (although this activation is weaker than that stimulated by microtubules). This suggests that the association of SH3-containing signaling proteins with dynamin may regulate endocytosis.

***Role of SH3 in Recruiting Specific Substrates to the Enzyme***

Besides SH2-mediated interactions, nonreceptor tyrosine kinases may also use SH3-mediated interactions to target themselves to some of their substrates. For instance, the Abl tyrosine kinase has binding sites selective for the SH3 domains of the Crk, Grb2, and Nck adaptor molecules. It is only upon binding to Abl that Crk becomes phosphorylated (Ren et al., 1994; Feller et al., 1994). The SH3 domain of Src family tyrosine kinases may target Src to its substrates: AFAP-110 and p68 (Flynn et al., 1993; Fumagalli et al., 1994; Taylor and Shalloway, 1994). Stable binding of these substrates to Src requires both the Src SH3 and the SH2 domains. This has led to the suggestion that binding of Src to these substrates is initiated by SH3-mediated interactions, followed by Tyr-phosphorylation, and ends with stable complex formation that is mediated by both SH3 and SH2 interactions. The stable association of tyrosine kinases with their substrates through noncatalytic domains may block the access of other substrates to the kinase. Signaling by tyrosine kinases may therefore involve pathways that require phosphorylation of only stoichiometric amounts of substrates.

***Role of SH3 in Interacting with Viral Proteins***

Because SH3-mediated interactions are critical for many cellular processes, it is no wonder that many viruses have pirated modified forms of SH3-containing proteins for their own devices (e.g., the Abelson murine leukemia virus). However, some viruses have evolved novel proteins to tap into tyrosine kinase signaling pathways. Recently, the HIV Nef protein was found to have a conserved PXXP motif that mediates Nef interaction with the SH3 domains of Hck and Lyn (Saksela et al., 1995). The PXXP motif is required for the higher replicative potential of Nef-bearing viruses, suggesting that the virus uses the SH3-mediated interaction to regulate a signaling pathway that facilitates viral growth.

**PH Domains**

The PH domain is a region of approximately 100 amino acids found in a wide variety of signaling and cytoskeletal proteins. The pleckstrin protein, the major protein kinase C substrate in platelets, has two such domains and was the first protein in which the domain was recognized (Haslam et al., 1993; Mayer et al., 1993). The sequence homology of the PH domain among proteins is generally low (Musacchio et al., 1993). The structure of three PH domains—the N-terminal PH domain and those from spectrin and dynamin—have been solved by NMR analysis (Downing et al., 1994; Macias et al., 1994; Yoon et al., 1994; Ferguson et al., 1994). Despite the low sequence homol-

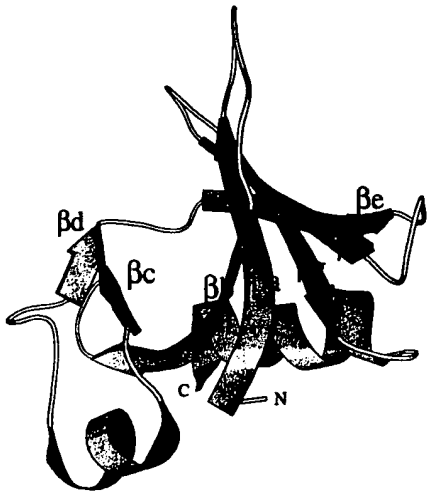



Figure 7. Schematic Diagram of the Spectrin PH Domain

This figure previously appeared in the paper by Gibson et al. (1994) and is reproduced by permission of the authors and publisher.

ogy among these domains, the peptide fold seen in these structures is virtually the same and provides structural evidence that the PH domain is a real functional unit. The basic PH fold contains two antiparallel β sheets and a long C-terminal α helix (Figure 7). The PH structures differ most in the loop regions between β strands. In particular, the CD loop tolerates large insertions (Gibson et al., 1994). Interestingly, for the PH domain of PLC-γ, the protein's two SH2 domains and one SH3 domain are predicted to insert into this loop. Many of the conserved hydrophobic residues found in PH domains are located in the interior of the protein (in contrast with the SH3 domain, where they form the aromatic-rich PPII-binding pocket), while the highly conserved charged residues are solvent-exposed. On one side of the PH domain there are a narrow cleft, three variable loops, and clusters of positive charges. This side of the molecule has been hypothesized to be the site of ligand binding (Gibson et al., 1994).

The function of PH domains is not known, although the presence of PH domains in a wide variety of proteins suggests that, like SH2 and SH3 domains, they mediate protein–protein interactions. Several PH domains bind to the βγ subunits of heterotrimeric G proteins. However, these interactions only require the C-terminal portion of the PH domain and also require residues that lie outside the PH domain, so it is not clear whether these interactions can be generalized for other PH domains. Mutations in the N-terminal portion of the PH domain found in the nonreceptor tyrosine kinase Btk result in the immune deficiency disorder agammaglobulinemia and illustrate the importance of this region of the PH domain (Gibson et al., 1994). Recently, it was suggested that PH domains might not serve as mediators of protein–protein interactions, but rather as mediators of protein–lipid interactions. Using a centrifugation assay, Harlan et al. (1994) reported that a number of PH domains associate specifically with lipid vesicles doped with $PtdInsP_2$. This interaction involves the previously mentioned positively charged area on the PH domain. The generality and biological relevance of PH–$PtdInsP_2$ interactions—and whether PH domains display other binding functions—remains to be determined.

## Perspectives

We have now reached the point in our understanding of these three modular domains where their occurrence and their structures are well characterized. For SH2 and SH3, but not PH, binding sites have been well delineated, and the structural basis of recognition and specificity is moderately clear. However, many questions remain open. These have been highlighted above and will not be reiterated. The issues for PH are much more open than for the others, because its binding specificity is not yet apparent. It is the most variable of the three in sequence; it may not have a stereotyped core-binding activity and could depend more on three-dimensional protein structure for its binding specificity. It should be remembered that the prototype binding protein, the immunoglobulin, binds to either linear epitopes or ones composed of segments from various parts of the linear structure, and there is nothing in common among the sites of binding.

There is sure to be much excitement in the future about the roles that these domains play in particular biological processes. Having already seen that they can mediate dimerization, bring proteins to their substrates at membranes, bring substrates to enzymes, modify the activities of enzymes, and aggregate subunits of multiprotein complexes, we can expect new classes of functions to emerge. Ultimately, the recognition of modular binding domains will take its historic place as one of the key events that will have allowed biologists to comprehend the full variety of mechanisms inherent in the signal transduction pathways of cells.

After submission of this review, there were three other reports of high-resolution structure determinations of SH3 domains bound to class II ligands: Lim et al. (1994), Terasawa et al. (1994), and Goudreau et al. (1994). These reports also demonstrate that class I and II ligands bind to SH3 domains in opposite orientations. In addition, a solution structure for the PH domain of dynamin has been reported: Fushman et al. (1995).

### Acknowledgments

Correspondence should be addressed to D. B. We would like to thank Martyn Botfield, John Kuriyan, Daniel Oprian, Matti Saraste, Joseph Schlessinger, Stuart Schreiber, Steven Shoelson, and Michael Waterfield for helpful discussions on this manuscript and for communicating results that are in press. We also thank Chi-Hon Lee, John Kuriyan, Andrea Musacchio, and Matti Saraste for redrawing figures for this manuscript, and *Structure*, *Trends in Biochemical Sciences*, and *Nature Structural Biology*, for permission to use previously published figures. We are also grateful to members of the Baltimore laboratory, especially Liz Alcamo and Marty Scott for comments on this manuscript. This work was supported by postdoctoral fellowships from the Helen Hay Whitney Foundation (G. B. C.) and the Cancer Research Institute (R. R.), and National Institutes of Health grant 5RO1 CA51462-06 (D. B.). Space limitation has precluded citation of all primary references, which are included in reviews cited.

**References**

Aderem, A. (1992). Signal transduction and the actin cytoskeleton: the roles of MARCKS and profilin. Trends Biochem. Sci. *17*, 438–443.

Adzhubei, A. A., and Sternberg, M. J. E. (1993). Left-handed polyproline II helices commonly occur in globular proteins. J. Mol. Biol. *229*, 472–493.

Bar-Sagi, D., Rotin, D., Batzer, A., Mandiyan, V., and Schlessinger, J. (1993). SH3 domains direct cellular localization of signaling molecules. Cell *74*, 83–91.

Bibbins, K. B., Boeuf, H., and Varmus, H. E. (1993). Binding of the Src SH2 domain to phosphopeptides is determined by residues in both the SH2 domain and the phosphopeptides. Mol. Cell. Biol. *13*, 7278–7287.

Cantley, L. C., Auger, K. R., Carpenter, C., Duckworth, B., Graziani, A., Kapeller, R., and Soltoff, S. (1991). Oncogenes and signal transduction. Cell *64*, 281–302.

Carpenter, C. L., Auger, K. R., Chanudhuri, M., Yoakim, M., Schaffhausen, B., Shoelson, S., and Cantley, L. C. (1993). Phosphoinositide 3-kinase is activated by phosphopeptides that bind to the SH2 domains of the 85-kDa subunit. J. Biol. Chem. *268*, 9478–9483.

Case, R. D., Piccione, E., Wolf, G., Benett, A. M., Lechleider, R. J., Neel, B. G., and Shoelson, S. E. (1994). SH-PTP2/Syp SH2 domain binding specificity is defined by direct interactions with platelet-derived growth factor β–receptor, epidermal growth factor receptor, and insulin receptor substrate-1-derived phosphopeptides. J. Biol. Chem. *269*, 10467–10474.

Chen, J. K., Lane, W. S., Brauer, A. W., Tanaka, A., and Schreiber, S. L. (1993). Biased combinatorial libraries: novel ligands for the SH3 domain of phosphatidylinositol 3-kinase. J. Am. Chem. Soc. *115*, 12591–12592.

Cicchetti, P., Mayer, B. J., Thiel, G., and Baltimore, D. (1992). Identification of a protein that binds to the SH3 region of Abl and is similar to Bcr and GAP-rho. Science *257*, 803–806.

Cooper, J. A., and Kashishian, A. (1993). In vivo binding properties of SH2 domains from GTPase-activating protein and phosphatidylinositol 3-kinase. Mol. Cell. Biol. *13*, 1737–1745.

Darnell, J. E., Jr., Kerr, I. M., and Stark, G. R. (1994). Jak–STAT pathways and transcriptional activation in response to INFs and other extracellular signaling proteins. Science *264*, 1415–1421.

Dérijard, B., Hibi, M., Wu, I-H., Barrett, T., Su, B., Deng, T., Karin, M., and Davis, R. J. (1994). JNK1: a protein kinase stimulated by UV light and Ha-Ras that binds and phosphorylates the c-Jun activation domain. Cell *76*, 1025–1037.

Downing, A. K., Driscoll, P. C., Gout, I., Salim, K., Zvelebil, M. J., and Waterfield, M. D. (1994). Three-dimentional solution structure of the pleckstrin homology domain from dynamin. Curr. Biol. *4*, 884–891.

Eck, M. J., Shoelson, S. E., and Harrison, S. C. (1993). Recognition of a high affinity phosphotyrosyl peptide by the Src homology 2 domain of p56lck. Nature *362*, 87–91.

Eck, M. J., Atwell, S. K., Shoelson, S. E., and Harrison, S. C. (1994). Structure of the regulatory domains of the Src-family tyrosine kinase Lck. Nature *368*, 764–769.

Felder, S., Zhou, M., Hu, P., Urena, J., Ullrich, A., Chaudhuri, M., White, M., Shoelson, S. E., and Schlessinger, J. (1993). SH2 domains exhibit high-affinity binding to tyrosine-phosphorylated peptides yet also exhibit rapid dissociation and exchange. Mol. Cell. Biol. *13*, 1449–1455.

Feller, S. M., Knudsen, B., and Hanafusa, H. (1994). c-Abl kinase regulates the protein binding activity of c-Crk. EMBO J. *13*, 2341–2351.

Feng, S., Chen, J. K., Yu, H., Simon, J. A., and Schreiber, S. L. (1994). A general model for SH3–ligand interactions involving dual binding modes. Science *266*, 1241–1247.

Ferguson, K. M., Lemmon, M. A., Schlessinger, J., and Sigler, P. B. (1994). Crystal structure at 2.2 Å resolution of the pleckstrin homology domain from human dynamin. Cell *79*, 199–209.

Finan, P., Shimizu, Y., Gout, I., Hsuan, J., Truong, O., Butcher, P., Bennett, P., Waterfield, M. D., and Kellie, S. (1994). An SH3 domain and proline-rich sequence mediate an interaction between two components of the phagocyte NADPH oxidase complex. J. Biol. Chem. *269*, 13752–13755.

Flynn, D. C., Leu, T. H., Reynolds, A. B., and Parsons, J. T. (1993). Identification and sequence analysis of cDNAs encoding a 110-kilodalton actin filament–associated pp60src substrate. Mol. Cell. Biol. *13*, 7892–7900.

Fumagalli, S., Totty, N. F., Hsuan, J. J., and Courtneidge, S. A. (1994). A target for Src in mitosis. Nature *368*, 871–874.

Fushman, D., Cahill, S., Lemmon, M. A., Schlessinger, J., and Cowburn, D. (1995). Solution structure of Pleckstrin homology domain of dynamin by heteronuclear NMR stectroscopy. Proc. Natl. Acad. Sci. *92*, in press.

Gibson, T. J., Hyvonen, M., Birney, E., Musacchio, A., and Saraste, M. (1994). PH domain—the first anniversary. Trends Biochem. Sci. *19*, 349–353.

Goudreau N., Cornille, F., Duchesne, M., Parker, F., Tocqué, B., Garbay, C., and Roques, B. P. (1994). NMR structure of the N-terminal SH3 domain of GRB2 and its complex with a proline-rich peptide from Sos. Nature Struct. Biol. *1*, 898–907.

Gout, I., Dhand, R., Hiles, I. D., Fry, M. J., Panayotou, G., Das, P., Truong, O., Totty, N. F., Hsuan, J., Booker, G. W., Campbell, I. D., and Waterfield, M. D. (1993). The GTPase dynamin binds to and is activated by a subset of SH3 domains. Cell *75*, 25–36.

Harlan, J. E., Hajduk, P. J., Yoon, H. S., and Fesik, S. W. (1994). Pleckstrin homology domains bind to phosphatidylinositol-4,5-bisphosphate. Nature *371*, 168–170.

Haslam, R., Kolde, H. B., and Hemmings, B. A. (1993). Pleckstrin domain homology. Nature *363*, 309–310.

Hu, P., and Schlessinger, J. (1994). Direct association of p110b phosphatidylinositol 3-kinase with p85 is mediated by an N-terminal fragment of p110b. Mol. Cell. Biol. *14*, 2577–2583.

Kazlauskas, A., Kashishian, A., Cooper, J. A., and Valius, M. (1992). GTPase-activating protein and phosphatidylinositol 3-kinase bind to distinct regions of the platelet-derived growth factor receptor β subunit. Mol. Cell. Biol. *12*, 2534–2544.

Klippel, A., Escobedo, J. A., Hirano, M., and Williams, L. T. (1994). The interaction of small domains between the subunits of phosphatidylinositol 3-kinase determines enzyme activity. Mol. Cell. Biol. *14*, 2675–2685.

Koch, C. A., Anderson, D., Moran, M. F., Ellis, C., and Pawson, T. (1991). SH2 and SH3 domains: elements that control interactions of cytoplasmic signaling proteins. Science *252*, 668–674.

Kyriakis, J. M., Banerjee, P., Nikolakaki, E., Dai, T., Rubie, E. A., Ahmad, M. F., Avruch, J., and Woodgett, J. R. (1994). The stress-activated protein kinase subfamily of c-Jun kinases. Nature *369*, 156–160.

Lechleider, R. J., Sugimoto, S., Bennett, A. M., Kashishian, A. S., Cooper, J. A., Shoelson, S. E., Walsh, C. T., and Neel, B. G. (1993). Activation of the SH2-containing phosphotyrosine phosphatase SH-PTP2 by its binding site, phosphotyrosine 1009, on the human platelet-derived growth factor receptor β. J. Biol. Chem. *268*, 21478–21481.

Lee, C., Kominos, D., Jacques, S., Margolis, B., Schlessinger, J., Shoelson, S. E., and Kuriyan, J. (1994). Crystal structures of peptide complexes of the amino-terminal SH2 domain of the Syp tyrosine phosphatase. Structure *2*, 423–438.

Lefkowitz, R. J. (1993). G protein–coupled receptor kinases. Cell *74*, 409–412.

Li, W., Nishimura, R., Kashishian, A., Batzer, A. G., Kim, W. J. H., Cooper, J. A., and Schlessinger, J. (1994). A new function for a phosphotyrosine phosphatase: linking Grb2–SOS to a receptor tyrosine kinase. Mol. Cell. Biol. *14*, 509–517.

Lim, W. A., Richards, F. M., and Fox, R. O. (1994). Structural determinants of peptide-binding orientation and of sequence specificity in SH3 domains. Nature *372*, 375–379.

Lisanti, M. P., Scherer, P. E., Tang, Z., and Sargiacomo, M. (1994). Caveolae, caveolin, and caveolin-rich membrane domains: a signalling hypothesis. Trends Cell Biol. *4*, 231–235.

Macias, M. J., Musacchio, A., Ponstingl, H., Nilges, M., Saraste, M., and Oschkinat, H. (1994). Structure of the pleckstrin homology domain from β-spectrin. Nature *368*, 675–677.

Marengere, L. E., Songyang, Z., Gish, G. D., Schaller, M. D., Parsons, J. T., Stern, M. J., Cantley, L. C., and Pawson, T. (1994). SH2 domain specificity and activity modified by a single residue. Nature *369*, 502–505.

Mayer, B., and Baltimore, D. (1993). Signalling through SH2 and SH3 domains. Trends Cell Biol. *3*, 8–13.

Mayer, B. J., Ren, R., Clark, K. L., and Baltimore, D. (1993). A putative modular domain present in diverse signaling proteins. Cell *73*, 629–630.

Miller, C. (1994). Potassium selectivity in proteins: oxygen cage or in the face? Science *261*, 1692–1693.

Mori, S., Rönnstrand, L., Yokote, K., Engström, A., Courtneidge, S. A., Claesson-Welsh, L., and Heldin, C. H. (1993). Identification of two juxtamembrane autophosphorylation sites in the PDGF β-receptor; involvement in the interaction with Src family tyrosine kinases. EMBO J. *12*, 2257–2264.

Muller, A. J., Pendergast, A.-M., Havlik, M. H., Puil, L., Pawson, T., and Witte, O. N. (1992). A limited set of SH2 domains binds BCR through a high-affinity phosphotyrosine-independent interaction. Mol. Cell. Biol. *12*, 5087–5093.

Musacchio, A., Gibson, T., Rice, P., Thompson, J., and Saraste, M. (1993). The PH domain: a common piece in the structural patchwork of signalling proteins. Trends Biochem. Sci. *18*, 343–348.

Musacchio, A., Saraste, M., and Wilmanns, M. (1994a). High-resolution crystal structures of tyrosine kinase SH3 domains complexed with proline-rich peptides. Nature Struct. Biol. *1*, 546–551.

Musacchio, A., Wilmanns, M., and Saraste, M. (1994b). Structure and function of the SH3 domain. Prog. Biophys. Mol. Biol. *61*, 283–297.

Myers, M. G. J., Sun, X. J., and White, M. F. (1994). The IRS-1 signaling system. Trends Biochem. Sci. *19*, 289–293.

Nishimura, R., Li, W., Kashishian, A., Mondino, A., Zhou, M., Cooper, J., and Schlessinger, J. (1993). Two signaling molecules share a phosphotyrosine-containing binding site in the platelet-derived growth factor receptor. Mol. Cell. Biol. *13*, 6889–6896.

Panayotou, G., Gish, G., End, P., Truong, O., Gout, I., Dhand, R., Fry, M. J., Hiles, I., Pawson, T., and Waterfield, M. D. (1993). Interactions between SH2 domains and tyrosine-phosphorylated platelet-derived growth factor beta-receptor sequences: analysis of kinetic parameters by a novel biosensor-based approach. Mol. Cell. Biol. *13*, 3567–3576.

Pascal, S. M., Singer, A. U., Gish, G., Yamazaki, T., Shoelson, S. E., Pawson, T., Kay, L. E., and Forman, K. J. (1994). Nuclear magnetic resonance structure of an SH2 domain of phospholipase C-γ1 complexed with a high affinity binding peptide. Cell *77*, 461–472.

Pawson, T., and Schlessinger, J. (1993). SH2 and SH3 domains. Curr. Biol. *3*, 434–442.

Piccione, E., Case, R. D., Domchek, S. M., Hu, P., Chaudhuri, M., Backer, J. M., Schlessinger, J., and Shoelson, S. E. (1993). Phosphatidylinositol 3-kinase p85 SH2 domain specificity defined by direct phosphopeptide/SH2 domain binding. Biochemistry *32*, 3197–3202.

Ren, R., Mayer, B. J., Cicchetti, P., and Baltimore, D. (1993). Identification of a 10-amino acid proline-rich SH3 binding site. Science *259*, 1157–1161.

Ren, R., Ye, Z., and Baltimore, D. (1994). Abl protein-tyrosine kinase selects the Crk adapter as a substrate using SH3-binding sites. Genes Dev. *8*, 783–795.

Rickles, R. J., Botfield, M. C., Weng, Z., Taylor, J. A., Green, O. M., Brugge, J. S., and Zoller, M. J. (1994). Identification of Src, Fyn, Lyn, PI3K, and Abl SH3 domain ligands using phage display libraries. EMBO J. *13*, 5598–5604.

Rotin, D., Bar-Sagi, D., O'Brodovich, H., Merilainen, J., Lehto, P. V., Canessa, C. M., Rossier, B. C., and Downey, G. P. (1994). A SH3 binding region in the epithelial $Na^+$ channel mediates its localization at the apical membrane. EMBO J. *13*, 4440–4450.

Rozakis-Adcock, M., Fernley, R., Wade, J., Pawson, T., and Bowtell, D. (1993). The SH2 and SH3 domains of mammalian Grb2 couple the EGF receptor to the Ras activator mSos1. Nature *363*, 83–85.

Saksela, K., Cheng, G., and Baltimore, D. (1995). Proline-rich (PxxP) motifs in HIV-1 Nef bind to SH3 domains of a subset of Src kinases and are required for the enhanced growth of $Nef^+$ viruses but not for downregulation of CD4. EMBO J. *14*, in press.

Seidel-Dugan, C., Meyer, B. E., Thomas, S. M., and Brugge, J. S. (1992). Effects of SH2 and SH3 deletions on the functional activities of wild-type and transforming variants of c-Src. Mol. Cell. Biol. *12*, 1835–1845.

Shoelson, S. E., Chatterjee, S., Chaudhuri, M., and White, M. F. (1992). YMXM motifs of IRS-1 define substrate specificity of the insulin receptor kinase. Proc. Natl. Acad. Sci. USA *89*, 2027–2031.

Shoelson, S. E., Sivaraja, M., Williams, K. P., Hu, P., Schlessinger, J., and Weiss, M. A. (1993). Specific phosphopeptide binding regulates a conformational change in the PI 3-kinase SH2 domain associated with enzyme activation. EMBO J. *12*, 795–802.

Songyang, Z., Shoelson, S. E., Chaudhuri, M., Gish, G., Pawson, T., Haser, W. G., King, F., Roberts, T., Ratnofsky, S., Lechleider, R. J., et al. (1993). SH2 domains recognize specific phosphopeptide sequences. Cell *72*, 767–778.

Songyang, Z., Shoelson, S. E., McGlade, J., Olivier, P., Pawson, T., Bustelo, X. R., Barbacid, M., Sabe, H., Hanafusa, H., Yi, T., et al. (1994). Specific motifs recognized by the SH2 domains of Csk, 3BP2, fps/fes, GRB-2, HCP, SHC, Syk, and Vav. Mol. Cell. Biol. *14*, 2777–2785.

Sugimoto, S., Wandless, T. J., Shoelson, S. E., Neel, B. G., and Walsh, C. T. (1994). Activation of the SH2-containing protein tyrosine phosphatase, SH-PTP2, by phosphotyrosine-containing peptides derived from insulin receptor substrate-1. J. Biol. Chem. *269*, 13614–13622.

Sumimoto, H., Kage, Y., Nunoi, H., Sasaki, H., Nose, T., Fukumaki, Y., Ohno, M., Minakami, S., and Takeshige, K. (1994). Role of Src homology 3 domains in assembly and activation of the phagocyte NADPH oxidase. Proc. Natl. Acad. Sci. USA *91*, 5345–5349.

Superti-Furga, G., Fumagalli, S., Koegl, M., Courtneidge, S. A., and Draetta, G. (1993). Csk inhibition of c-Src activity requires both the SH2 and SH3 domains of Src. EMBO J. *12*, 2625–2634.

Taylor, S. J., and Shalloway, D. (1994). An RNA-binding protein associated with Src through its SH2 and SH3 domains in mitosis. Nature *368*, 867–871.

Terasawa, H., Kohda, D., Hatanaka, H., Tsuchiya, S., Ogura, K., Nagata, K., Ishii, S., Mandiyan, V., Ullrich, A., Schlessinger, J., and Inagaki, F. (1994). Nature Struct. Biol. *1*, 891–897.

Valius, M., and Kazlauskas, A. (1993). Phospholipase C-γ1 and phosphatidylinositol 3 kinase are the downstream mediators of the PDGF receptor's mitogenic signal. Cell *73*, 321–334.

Waksman, G., Kominos, D., Robertson, S. C., Pant, N., Baltimore, D., Birge, R. B., Cowburn, D., Hanafusa, H., Mayer, B. J., Overduin, M., Resh, M. D., Rios, C. B., Silverman, L., and Kuriyan, J. (1992). Crystal structure of the phosphotyrosine recognition domain (SH2) of the v-*src* tyrosine kinase complexed with tyrosine phosphorylated peptides. Nature *358*, 646–653.

Waksman, G., Shoelson, S. E., Pant, N., Cowburn, D., and Kuriyan, J. (1993). Binding of a high affinity phosphotyrosyl peptide to the Src SH2 domain: crystal structures of the complexed and peptide-free forms. Cell *72*, 779–790.

Ye, Z.-S., and Baltimore, D. (1995). Binding of Vav to Grb2 through dimerization of Src homology 3 domains. Proc. Natl. Acad. Sci. USA *92*, in press.

Yoon, H. S., Hajduk, P. J., Petros, A. M., Olejniczak, E. T., Meadows, R. P., and Fesil, S. W. (1994). Solution structure of a pleckstrin-homology domain. Nature *369*, 672–675.

Yu, H., Chen, J. K., Feng, S., Dalgarno, D. C., Brauer, A. W., and Schreiber, S. L. (1994). Structural basis for the binding of proline-rich peptides to SH3 domains. Cell *76*, 933–945.

# This Page is Inserted by IFW Indexing and Scanning Operations and is not part of the Official Record

## BEST AVAILABLE IMAGES

Defective images within this document are accurate representations of the original documents submitted by the applicant.

Defects in the images include but are not limited to the items checked:

- ☑ **BLACK BORDERS**
- ☐ **IMAGE CUT OFF AT TOP, BOTTOM OR SIDES**
- ☐ **FADED TEXT OR DRAWING**
- ☐ **BLURRED OR ILLEGIBLE TEXT OR DRAWING**
- ☐ **SKEWED/SLANTED IMAGES**
- ☐ **COLOR OR BLACK AND WHITE PHOTOGRAPHS**
- ☐ **GRAY SCALE DOCUMENTS**
- ☐ **LINES OR MARKS ON ORIGINAL DOCUMENT**
- ☐ **REFERENCE(S) OR EXHIBIT(S) SUBMITTED ARE POOR QUALITY**
- ☐ **OTHER:** _____

**IMAGES ARE BEST AVAILABLE COPY.**
**As rescanning these documents will not correct the image problems checked, please do not report these problems to the IFW Image Problem Mailbox.**